AQ

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 May 2001 (03.05.2001)

PCT

(10) International Publication Number
WO 01/30991 A3

(51) International Patent Classification[7]: C07K 14/705, C12Q 1/68

(21) International Application Number: PCT/US00/23021

(22) International Filing Date: 22 August 2000 (22.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/426,290 25 October 1999 (25.10.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/426,290 (CIP)
Filed on 25 October 1999 (25.10.1999)

(71) Applicant *(for all designated States except US)*: **DECODE GENETICS EHF.** [IS/IS]: Lynghals 1, IS-110 Reykjavik (IS).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **OLAFSDOTTIR, Berglind, Ran** [IS/IS]: Eskihlid 15, IS-105 Reykjavik (IS). **GULCHER, Jeffrey** [US/US]: Unit M, 130 South Canal Street, Chicago, IL 60606 (US).

(74) Agents: **CARROLL, Alice, O.** et al.: Hamilton, Brook, Smith & Reynolds, P.C., 530 Virginia Road, P.O. Box 9133, Concord, MA 01742-9133 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report: 6 December 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/30991 A3

(54) Title: HUMAN NARCOLEPSY GENE

(57) Abstract: The gene for hypocretin (orexin) receptor 2 (HCRTR2), which is associated with narcolepsy, is disclosed. Also described are methods of diagnosis of narcolepsy, pharmaceutical compositions comprising nucleic acids comprising the HCRTR2 gene, as well as methods of therapy of narcolepsy.

INTERNATIONAL SEARCH REPORT

Internat. Application No
PCT/US 00/23021

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07K14/705 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, CHEM ABS Data, WPI Data, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 96 34877 A (HUMAN GENOME SCIENCES INC.; LI YI (US); ROSEN CRAIG A (US); SOPPET) 7 November 1996 (1996-11-07) the whole document ---	1-7
Y	LIN LING ET AL: "The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene" CELL,CELL PRESS, CAMBRIDGE, NA,US, vol. 98, no. 3, 6 August 1999 (1999-08-06), pages 365-376, XP002153571 ISSN: 0092-8674 abstract; figure 6 ---	1-7

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 March 2001	19. 04. 01

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hardon, E

1

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	SAKURAI T ET AL: "Oxerins and oxerin receptors: A family of hypothalamic neuropeptides and G Protein-coupled receptors that regulate feeding behaviour" CELL,CELL PRESS, CAMBRIDGE, NA,US, vol. 92, 20 February 1998 (1998-02-20), pages 573-585, XP002105412 ISSN: 0092-8674 page 585, column 2; figure 2 ---	1-7
Y	ALDRICH, MICHAEL S. ET AL: "Narcolepsy and the hypocretin receptor 2 gene" NEURON (1999), 23(4), 625-626 , 1999, XP000973742 the whole document ---	1-7
Y	SIEGEL, JEROME M.: "Narcolepsy: A key role for hypocretins (orexins)" CELL (CAMBRIDGE, MASS.) (1999), 98(4), 409-412 , 20 August 1999 (1999-08-20), XP000941943 the whole document ---	1-7
A	LECEA L ET AL: "The hypocretins: hypothalamus-specific peptides with neuroexcitatory activity" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA,NATIONAL ACADEMY OF SCIENCE. WASHINGTON,US, vol. 95, January 1998 (1998-01), pages 322-327, XP002105411 ISSN: 0027-8424 the whole document ---	1-7
T	PEYRON CHRISTELLE ET AL: "A mutation in a case of early onset narcolepsy and a generalized absence of hypocretin peptides in human narcoleptic brains" NATURE MEDICINE,NATURE PUBLISHING, CO,US, vol. 6, no. 9, September 2000 (2000-09), pages 991-997, XP002153570 ISSN: 1078-8956 -----	1-7

1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 00/23021

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 7
because they relate to subject matter not required to be searched by this Authority, namely:
Rule 39.1(iv) PCT - Method for treatment of the human or animal body by therapy

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Internat. l Application No
PCT/US 00/23021

Patent document cited in search report		Publication date	Patent family member(s)			Publication date
WO 9634877	A	07-11-1996	CA	2220036	A	07-11-1996
			AU	715286	B	20-01-2000
			AU	2470795	A	21-11-1996
			EP	0828751	A	18-03-1998
			JP	11505110	T	18-05-1999

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 May 2001 (03.05.2001)

PCT

(10) International Publication Number
WO 01/30991 A2

(51) International Patent Classification[7]: C12N 15/00

(21) International Application Number: PCT/US00/23021

(22) International Filing Date: 22 August 2000 (22.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/426,290 25 October 1999 (25.10.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/426,290 (CIP)
Filed on 25 October 1999 (25.10.1999)

(71) Applicant *(for all designated States except US)*: **DECODE GENETICS EHF.** [IS/IS]; Lynghals 1, IS-110 Reykjavik (IS).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **OLAFSDOTTIR, Berglind, Ran** [IS/IS]; Eskihlid 15, IS-105 Reykjavik (IS). **GULCHER, Jeffrey** [US/US]; Unit M, 130 South Canal Street, Chicago, IL 60606 (US).

(74) Agents: **CARROLL, Alice, O.** et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/30991 A2

(54) Title: HUMAN NARCOLEPSY GENE

(57) Abstract: The gene for hypocretin (orexin) receptor 2 (HCRTR2), which is associated with narcolepsy, is disclosed. Also described are methods of diagnosis of narcolepsy, pharmaceutical compositions comprising nucleic acids comprising the HCRTR2 gene, as well as methods of therapy of narcolepsy.

HUMAN NARCOLEPSY GENE

RELATED APPLICATION

This application is a Continuation-in-Part of U.S. Serial No. 09/426,290, filed October 25, 1999, the entire teachings of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

Narcolepsy, a disorder which affects approximately 1 in 2,000 individuals, is characterized by daytime sleepiness, sleep fragmentation, and symptoms of abnormal rapid eye movement (REM) sleep that include cataplexy (loss of muscle tone), sleep paralysis, and hypnagogic hallucinations (Aldrich, M.S., *Neurology 42*:34-43 (1992); Siegel, J.M., *Cell 98*:409-412 (1999)). In humans, susceptibility to narcolepsy has been associated with a specific human leukocyte antigen (HLA) alleles, including DQB1*0602 (Mignot, E., *Neurology 50*:S16-22 (1998); Kadotani, H. *et al., Genome Res. 8*:427-434 (1998); Faraco, J. *et al., J. Hered. 90*:129-132 (1999)); however, attempts to verify narcolepsy as an autoimmune disorder have failed (Mignot, E. *et al., Adv. Neuroimmunol. 5*:23-37 (1995); Mignot, E., *Curr. Opin. Pulm. Med. 2*:482-487 (1996)). In a canine model of narcolepsy, the disorder is transmitted as an autosomal recessive trait, *canarc-1* (Foutz, A.S. *et al., Sleep 1*:413-421 91979); Baker, T.L. and Dement, W.C., Brain Mechanisms of Sleep (D.J. McGinty *et al., eds.*, New York: Raven Press, pp. 199-233 (1985)). The possibility of linkage between *canarc-1* and the canine major histocompatibility complex has been excluded (Mignot, E. *et al., Proc. Natl. Acad. Sci. USA 88*:3475-3478 (1991)).

A mutation in the hypocretin (orexin) receptor 2 gene in canines has been identified in narcolepsy (Lin, L. *et al., Cell 98*:365-376 (1999)); Hypocrexins/orexins (orexin-A and -B) are neuropeptides associated with regulation of food consumption (de Lecea, L., *et al., Proc. natl. Acad. Sci. USA 95*:322-327 (1998); Sakurai, T. *et al., Cell 92*:573-585 (1998)) as well as other possible functions (Peyron, C. *et al., J. Neurosci. 18*:9996-10015 (1998)). Human cDNA of receptors for orexins have been cloned (Sakurai, T. *et al., Cell 92*:573-585 (1998)), however, full human genes for the orexin receptors have not yet been identified.

Diagnosis of narcolepsy is difficult, as it is necessary to distinguish narcolepsy from other conditions such as chronic fatigue syndrome or other sleep disorders (Ambrogetti, A. and Olson, L.C., *Med. J. Aust. 160*:426-429 (1994); Aldrich, M.S., *Neurology 50*:S2-7 (1998)). Methods of diagnosing narcolepsy based on specific criteria would facilitate identification of the disease, reduce the time and expense associated with diagnosis, and expedite commencement of treatment.

SUMMARY OF THE INVENTION

As described herein, a full gene for the human hypocretin (orexin) receptor 2 (HCRTR2) has been identified. The sequence of the HCRTR2 gene as described herein is shown in Figure 1 (SEQ ID NO: 1). Accordingly, this invention pertains to an isolated nucleic acid molecule containing the HCRTR2 gene. The invention also relates to DNA constructs comprising the nucleic acid molecules described herein operatively linked to a regulatory sequence, and to recombinant host cells, such as bacterial cells, fungal cells, plant cells, insect cells and mammalian cells, comprising the nucleic acid molecules described herein operatively linked to a regulatory sequence. The invention also pertains to methods of diagnosing narcolepsy in an individual. The methods include detecting the presence of a mutation in the HCRTR2 gene. The invention additionally pertains to pharmaceutical compositions comprising the HCRTR2 nucleic acids of the invention. The invention further pertains to methods of treating narcolepsy, by administering HCRTR2 nucleic acids

of the invention or compositions comprising the HCRTR2 nucleic acids. The methods of the invention allow the accurate diagnosis of narcolepsy and reduce the need for time-consuming and expensive sleep laboratory assessments.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1A to Fig. 1AY depict the sequence of the human orexin receptor 2 gene (SEQ ID NO:1) and the encoded receptor (SEQ ID NO:2).

The foregoing and other objects, features and advantages of the invention will be apparent from the following more particular description of preferred embodiments of the invention, as illustrated in the accompanying drawings

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to a human hypocretin (orexin) receptor 2 (HCRTR2) gene, and the relationship of the gene to narcolepsy. As described herein, Applicants have isolated the HCRTR2 gene. The gene and its products are implicated in the pathogenesis of narcolepsy, as mutations in a closely related receptor, hypocretin (orexin) receptor 2, have been associated with the presence of narcolepsy in a well-established canine model of narcolepsy (Lin, L. *et al., Cell 98*:365-376 (1999)) .

NUCLEIC ACIDS OF THE INVENTION

Accordingly, the invention pertains to an isolated nucleic acid molecule containing the human HCRTR2 gene. The term, "HCRTR2 gene," refers to an isolated genomic nucleic acid molecule that encodes the human hypocretin (orexin) receptor 2. As used herein, the term, "genomic nucleic acid molecule" indicates that the nucleic acid molecule contains introns and exons as are found in genomic DNA (i.e., not cDNA). The nucleic acid molecules can be double-stranded or single-stranded; single stranded nucleic acid molecules can be either the coding (sense) strand or the non-coding (antisense) strand. The nucleic acid molecule can additionally contain a marker sequence, for example, a nucleotide sequence which encodes a polypeptide, to assist in isolation or purification of the polypeptide. Such

sequences include, but are not limited to, those which encode a glutathione-S-transferase (GST) fusion protein and those which encode a hemagglutinin A (HA) peptide marker from influenza. In a preferred embodiment, the nucleic acid molecule has the sequence shown in the Figure (SEQ ID NO:1).

As used herein, an "isolated" or "substantially pure" gene or nucleic acid molecule is intended to mean a gene which is not flanked by nucleotide sequences which normally (in nature) flank the gene (as in other genomic sequences). Thus, an isolated gene can include a gene which is synthesized chemically or by recombinant means. Thus, recombinant DNA contained in a vector are included in the definition of "isolated" as used herein. Also, isolated nucleotide sequences include recombinant DNA molecules in heterologous host cells, as well as partially or substantially purified DNA molecules in solution. Such isolated nucleotide sequences are useful in the manufacture of the encoded protein, as probes for isolating homologous sequences (e.g., from other mammalian species), for gene mapping (e.g., by *in situ* hybridization with chromosomes), or for detecting expression of the HCRTR2 gene in tissue (e.g., human tissue), such as by Northern blot analysis.

The present invention also encompasses variations of the nucleic acid sequences of the invention. Such variations can be naturally-occurring, such as in the case of allelic variation, or non-naturally-occurring, such as those induced by various mutagens and mutagenic processes. Intended variations include, but are not limited to, addition, deletion and substitution of one or more nucleotides which can result in conservative or non-conservative amino acid changes, including additions and deletions. Preferably, the nucleotide or amino acid variations are silent or conserved; that is, they do not alter the characteristics or activity of the hypocretin (orexin) receptor 2.

Other alterations of the nucleic acid molecules of the invention can include, for example, labeling, methylation, internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, carbamates), charged linkages (e.g., phosphorothioates, phosphorodithioates), pendent moieties (e.g., polypeptides), intercalators (e.g., acridine, psoralen),

chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids). Also included are synthetic molecules that mimic nucleic acid molecules in the ability to bind to a designated sequences via hydrogen bonding and other chemical interactions. Such molecules include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

The invention also relates to fragments of the isolated nucleic acid molecules described herein. The term "fragment" is intended to encompass a portion of a nucleic acid sequence described herein which is from at least about 25 contiguous nucleotides to at least about 50 contiguous nucleotides or longer in length. One or more introns can also be present. Such fragments are useful as probes, e.g., for diagnostic methods, as described below and also as primers or probes. Particularly preferred primers and probes selectively hybridize to a nucleic acid molecule containing the HCRTR2 gene described herein.

The invention also pertains to nucleic acid molecules which hybridize under high stringency hybridization conditions, such as for selective hybridization, to a nucleotide sequence described herein (e.g., nucleic acid molecules which specifically hybridize to a nucleic acid containing the HCRTR2 gene described herein). Hybridization probes are oligonucleotides which bind in a base-specific manner to a complementary strand of nucleic acid. Suitable probes include polypeptide nucleic acids, as described in (Nielsen *et al.*, *Science* 254, 1497-1500 (1991)).

Such nucleic acid molecules can be detected and/or isolated by specific hybridization (e.g., under high stringency conditions). "Stringency conditions" for hybridization is a term of art which refers to the incubation and wash conditions, e.g., conditions of temperature and buffer concentration, which permit hybridization of a particular nucleic acid to a second nucleic acid; the first nucleic acid may be perfectly (i.e., 100%) complementary to the second, or the first and second may share some degree of complementarity which is less than perfect (e.g., 60%, 75%, 85%, 95%). For example, certain high stringency conditions can be used which distinguish perfectly complementary nucleic acids from those of less complementarity.

"High stringency conditions", "moderate stringency conditions" and "low stringency conditions" for nucleic acid hybridizations are explained on pages 2.10.1-2.10.16 and pages 6.3.1-6 in *Current Protocols in Molecular Biology* (Ausubel, F.M. *et al.*, "*Current Protocols in Molecular Biology*", John Wiley & Sons, (1998)) the teachings of which are hereby incorporated by reference. The exact conditions which determine the stringency of hybridization depend not only on ionic strength (e.g., 0.2XSSC, 0.1XSSC), temperature (e.g., room temperature, 42°C, 68°C) and the concentration of destabilizing agents such as formamide or denaturing agents such as SDS, but also on factors such as the length of the nucleic acid sequence, base composition, percent mismatch between hybridizing sequences and the frequency of occurrence of subsets of that sequence within other non-identical sequences. Thus, high, moderate or low stringency conditions can be determined empirically. By varying hybridization conditions from a level of stringency at which no hybridization occurs to a level at which hybridization is first observed, conditions which will allow a given sequence to hybridize (e.g., selectively) with the most similar sequences in the sample can be determined.

Exemplary conditions are described in Krause, M.H. and S.A. Aaronson, *Methods in Enzymology, 200*:546-556 (1991). Also, in, Ausubel, *et al.*, "*Current Protocols in Molecular Biology*", John Wiley & Sons, (1998), which describes the determination of washing conditions for moderate or low stringency conditions. Washing is the step in which conditions are usually set so as to determine a minimum level of complementarity of the hybrids. Generally, starting from the lowest temperature at which only homologous hybridization occurs, each °C by which the final wash temperature is reduced (holding SSC concentration constant) allows an increase by 1% in the maximum extent of mismatching among the sequences that hybridize. Generally, doubling the concentration of SSC results in an increase in T_m of ~17°C. Using these guidelines, the washing temperature can be determined empirically for high, moderate or low stringency, depending on the level of mismatch sought.

For example, a low stringency wash can comprise washing in a solution containing 0.2XSSC/0.1% SDS for 10 min at room temperature; a moderate

stringency wash can comprise washing in a prewarmed solution (42°C) solution containing 0.2XSSC/0.1% SDS for 15 min at 42°C; and a high stringency wash can comprise washing in prewarmed (68°C) solution containing 0.1XSSC/0.1%SDS for 15 min at 68°C. Furthermore, washes can be performed repeatedly or sequentially to obtain a desired result as known in the art. Equivalent conditions can be determined by varying one or more of the parameters given as an example, as known in the art, while maintaining a similar degree of identity or similarity between the target nucleic acid molecule and the primer or probe used.

Hybridizable nucleic acid molecules are useful as probes and primers, e.g., for diagnostic applications, as described below. As used herein, the term "primer" refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis under appropriate conditions (*e.g.*, in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer, but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template, but must be sufficiently complementary to hybridize with a template. The term "primer site" refers to the area of the target DNA to which a primer hybridizes. The term "primer pair" refers to a set of primers including a 5' (upstream) primer that hybridizes with the 5' end of the DNA sequence to be amplified and a 3' (downstream) primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

The invention also pertains to nucleotide sequences which have a substantial identity with the nucleotide sequences described herein; particularly preferred are nucleotide sequences which have at least about 70%, and more preferably at least about 80% identity, and even more preferably at least about 90% identity, with nucleotide sequences described herein. Particularly preferred in this instance are nucleotide sequences encoding hypocretin (orexin) receptor 2.

To determine the percent identity of two nucleotide sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first nucleotide sequence). The nucleotides at corresponding nucleotide positions are then compared. When a position in the first sequence is occupied by the same nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical positions/total # of positions x 100).

The determination of percent identity between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin *et al.* (*Proc. Natl. Acad. Sci. USA*, *90*:5873-5877 (1993)). Such an algorithm is incorporated into the NBLAST program which can be used to identify sequences having the desired identity to nucleotide sequences of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (*Nucleic Acids Res, 25*:3389-3402 (1997)). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., NBLAST) can be used. See http://www.ncbi.nlm.nih.gov. In one embodiment, parameters for sequence comparison can be set at W=12. Parameters can also be varied (e.g., W=5 or W=20). The value "W" determines how many continuous nucleotides must be identical for the program to identify two sequences as containing regions of identity.

The invention also provides expression vectors containing a nucleic acid comprising the HCRTR2 gene, operatively linked to at least one regulatory sequence. Many such vectors are commercially available, and other suitable vectors can be readily prepared by the skilled artisan. "Operatively linked" is intended to mean that the nucleic acid sequence is linked to a regulatory sequence in a manner which allows expression of the nucleic acid sequence. Regulatory sequences are art-recognized and are selected to produce a hypocretin (orexin) receptor 2. Accordingly, the term "regulatory sequence" includes promoters, enhancers, and

other expression control elements such as those described in Goeddel, *Gene Expression Technology: Methods in Enzymology 185*, Academic Press, San Diego, CA (1990). For example, the native regulatory sequences or regulatory sequences native to the transformed host cell can be employed. It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the receptor desired to be expressed. For instance, the gene of the present invention can be expressed by ligating the gene into a vector suitable for expression in either prokaryotic cells, eukaryotic cells or both (see, for example, Broach, *et al.*, *Experimental Manipulation of Gene Expression*, ed. M. Inouye (Academic Press, 1983) p. 83; *Molecular Cloning: A Laboratory Manual*, 2nd Ed., ed. Sambrook *et al.* (Cold Spring Harbor Laboratory Press, 1989) Chapters 16 and 17). Typically, expression constructs will contain one or more selectable markers, including, but not limited to, the gene that encodes dihydrofolate reductase and the genes that confer resistance to neomycin, tetracycline, ampicillin, chloramphenicol, kanamycin and streptomycin resistance. Vectors can also include, for example, an autonomously replicating sequence (ARS), expression control sequences, ribosome-binding sites, RNA splice sites, polyadenylation sites, transcriptional terminator sequences, secretion signals and mRNA stabilizing sequences.

Prokaryotic and eukaryotic host cells transformed by the described vectors are also provided by this invention. For instance, cells which can be transformed with the vectors of the present invention include, but are not limited to, bacterial cells such as *E. coli* (e.g., *E. coli* K12 strains), *Streptomyces*, *Pseudomonas*, *Serratia marcescens* and *Salmonella typhimurium*, insect cells (baculovirus), including *Drosophila*, fungal cells, such as yeast cells, plant cells and mammalian cells, such as thymocytes, Chinese hamster ovary cells (CHO), and COS cells. The host cells can be transformed by the described vectors by various methods (e.g., electroporation, transfection using calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances; microprojectile bombardment; lipofection, infection where the vector is an infectious agent such as a retroviral genome, and other methods), depending on the type of cellular host.

The nucleic acid molecules of the present invention can be produced, for example, by replication in a suitable host cell, as described above. Alternatively, the nucleic acid molecules can also be produced by chemical synthesis.

The nucleotide sequences of the nucleic acid molecules described herein (e.g., a nucleic acid molecule comprising SEQ ID NO:1) can be amplified by methods known in the art. For example, this can be accomplished by e.g., PCR. *See generally PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, *et al.*, Academic Press, San Diego, CA, 1990); Mattila *et al.*, *Nucleic Acids Res.* 19, 4967 (1991); Eckert *et al.*, *PCR Methods and Applications* 1, 17 (1991); *PCR* (eds. McPherson *et al.*, IRL Press, Oxford); and U.S. Patent 4,683,202.

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4, 560 (1989), Landegren *et al.*, *Science* 241, 1077 (1988), transcription amplification (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA* 86, 1173 (1989)), and self-sustained sequence replication (Guatelli *et al.*, *Proc. Nat. Acad. Sci. USA,* 87, 1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

The amplified DNA can be radiolabeled and used as a probe for screening a library or other suitable vector to identify homologous nucleotide sequences. Corresponding clones can be isolated, DNA can be obtained following *in vivo* excision, and the cloned insert can be sequenced in either or both orientations by art recognized methods, to identify the correct reading frame encoding a protein of the appropriate molecular weight. For example, the direct analysis of the nucleotide sequence of homologous nucleic acid molecules of the present invention can be accomplished using either the dideoxy chain termination method or the Maxam - Gilbert method (see Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd Ed., CSHP, New York 1989); Zyskind *et al.*, *Recombinant DNA Laboratory*

Manual, (Acad. Press, 1988)). Using these or similar methods, the protein(s) and the DNA encoding the protein can be isolated, sequenced and further characterized.

METHODS OF DIAGNOSIS

The nucleic acids and the proteins described above can be used to detect, in an individual, a mutation in the HCRTR2 gene that is associated with narcolepsy. In one embodiment of the invention, diagnosis of narcolepsy is made by detecting a mutation in the HCRTR2 gene. The mutation can be the insertion or deletion of a single nucleotide, or of more than one nucleotide, resulting in a frame shift mutation; the change of at least one nucleotide, resulting in a change in the encoded amino acid; the change of at least one nucleotide, resulting in the generation of a premature stop codon; the deletion of several nucleotides, resulting in a deletion of one or more amino acids encoded by the nucleotides; the insertion of one or several nucleotides, such as by unequal recombination or gene conversion, resulting in an interruption of the coding sequence of the gene; duplication of all or a part of the gene; transposition of all or a part of the gene; or rearrangement of all or a part of the gene. More than one such mutation may be present in a single gene. Such sequence changes cause a mutation in the receptor encoded by the HCRTR2 gene. For example, if the mutation is a frame shift mutation, the frame shift can result in a change in the encoded amino acids, and/or can result in the generation of a premature stop codon, causing generation of a truncated receptor. Alternatively, a mutation associated with narcolepsy can be a synonymous mutation in one or more nucleotides (i.e., a mutation that does not result in a change in the receptor encoded by the HCRTR2 gene, such as a mutation in an intron or an untranslated portion of the gene). Such a polymorphism may alter splicing sites, affect the stability or transport of mRNA, or otherwise affect the transcription or translation of the gene. A HCRTR2 gene that has any of the mutations described above is referred to herein as a "mutant gene." It is likely that a mutation in the HCRTR2 gene is associated with narcolepsy in humans because of the association between a mutation in the HCRTR2 gene and narcolepsy in dogs (Lin, L. *et al.*, *Cell 98*:365-376 (1999), the entire teachings of which are incorporated herein by reference). In a preferred

embodiment, the mutation in the HCRTR2 gene is to a deletion mutation, for example, a deletion that corresponds to the deletions found in the hypocretin (orexin) receptor 2 in narcoleptic dogs as described by Lin *et al.*, *supra* (e.g., a deletion of one or more exons, such as a deletion of the fourth exon, that can be caused by insertion of one or more nucleotides upstream of the splice site of the exon, or a deletion of exon 6, that can be caused by a G to A transition in the splice junction consensus sequence). In another preferred embodiment, the mutation in the HCRTR2 gene is mutation that effects a "knockout" of the entire gene, such as deletion of the first exon as described by Chemelli, R.M. *et al*, (*Cell 98*:437-451 (1999), the entire teachings of which are incorporated herein). In a third preferred embodiment, the mutation in the HCRTR2 gene is a mutation in an intron, that affects splicing (joining of exons) during translation of the HCRTR2 gene.

In a first method of diagnosing narcolepsy, hybridization methods, such as Southern analysis, are used (see Current Protocols in Molecular Biology, Ausubel, F. *et al.*, eds., John Wiley & Sons, including all supplements through 1999). For example, a test sample of genomic DNA, RNA, or cDNA, is obtained from an individual suspected of having (or carrying a defect for) narcolepsy (the "test individual"). The individual can be an adult, child, or fetus. The test sample can be from any source which contains genomic DNA, such as a blood sample, sample of amniotic fluid, sample of cerebrospinal fluid, or tissue sample from skin, muscle, placenta, gastrointestinal tract or other organs. A test sample of DNA from fetal cells or tissue can be obtained by appropriate methods, such as by amniocentesis or chorionic villus sampling. The DNA, RNA, or cDNA sample is then examined to determine whether a mutation in the HCRTR2 gene is present. The presence of the mutation can be indicated by hybridization of the gene in the test sample to a nucleic acid probe. A "nucleic acid probe", as used herein, can be a DNA probe or an RNA probe; the nucleic acid probe contains at least one mutation in the HCRTR2 gene. The probe can be one of the nucleic acid molecules described above (e.g., the gene, a vector comprising the gene, etc.)

To diagnose narcolepsy by hybridization, a hybridization sample is formed by contacting the test sample containing a HCRTR2 gene, with at least one nucleic

acid probe. The hybridization sample is maintained under conditions which are sufficient to allow specific hybridization of the nucleic acid probe to the HCRTR2 gene. "Specific hybridization", as used herein, indicates exact hybridization (e.g., with no mismatches). Specific hybridization can be performed under high stringency conditions or moderate stringency conditions, for example, as described above. In a particularly preferred embodiment, the hybridization conditions for specific hybridization are high stringency.

Specific hybridization, if present, is then detected using standard methods. If specific hybridization occurs between the nucleic acid probe and the HCRTR2 gene in the test sample, then the HCRTR2 gene has the mutation that is present in the nucleic acid probe. More than one nucleic acid probe can also be used concurrently in this method. Specific hybridization of any one of the nucleic acid probes is indicative of a mutation in the HCRTR2 gene, and is therefore diagnostic for narcolepsy.

In another hybridization method, Northern analysis (see Current Protocols in Molecular Biology, Ausubel, F. *et al.*, eds., John Wiley & Sons, *supra*) is used to identify the presence of a mutation associated with narcolepsy. For Northern analysis, a test sample of RNA is obtained from the individual by appropriate means. Specific hybridization of a nucleic acid probe, as described above, to RNA from the individual is indicative of a mutation in the HCRTR2 gene, and is therefore diagnostic for narcolepsy

For representative examples of use of nucleic acid probes, see, for example, U.S. Patents No. 5,288,611 and 4,851,330. Alternatively, a peptide nucleic acid (PNA) probe can be used instead of a nucleic acid probe in the hybridization methods described above. PNA is a DNA mimic having a peptide-like, inorganic backbone, such as N-(2-aminoethyl)glycine units, with an organic base (A, G, C, T or U) attached to the glycine nitrogen via a methylene carbonyl linker (see, for example, Nielsen, P.E. *et al., Bioconjugate Chemistry*, 1994, 5, American Chemical Society, p. 1 (1994). The PNA probe can be designed to specifically hybridize to a gene having a polymorphism associated with autoimmune disease. Hybridization of the PNA probe to the HCRTR2 gene is diagnostic for narcolepsy..

In another method of the invention, mutation analysis by restriction digestion can be used to detect mutant genes, or genes containing polymorphisms, if the mutation or polymorphism in the gene results in the creation or elimination of a restriction site. A test sample containing genomic DNA is obtained from the individual. Polymerase chain reaction (PCR) can be used to amplify the HCRTR2 gene (and, if necessary, the flanking sequences) in the test sample of genomic DNA from the test individual. RFLP analysis is conducted as described *(see* Current Protocols in Molecular Biology, supra). The digestion pattern of the relevant DNA fragment indicates the presence or absence of the mutation in the HCRTR2 gene, and therefore indicates the presence or absence of narcolepsy.

Sequence analysis can also be used to detect specific mutations in the HCRTR2 gene. A test sample of DNA is obtained from the test individual. PCR can be used to amplify the gene, and/or its flanking sequences. The sequence of the HCRTR2 gene, or a fragment of the gene is determined, using standard methods. The sequence of the gene (or gene fragment) is compared with the nucleic acid sequence of the gene, as described above. The presence of a mutation in the HCRTR2 gene indicates that the individual has narcolepsy.

Allele-specific oligonucleotides can also be used to detect the presence of a mutation in the HCRTR2 gene, through the use of dot-blot hybridization of amplified proteins with allele-specific oligonucleotide (ASO) probes (see, for example, Saiki, R. *et al.*, (1986), *Nature (London) 324*:163-166). An "allele-specific oligonucleotide" (also referred to herein as an "allele-specific oligonucleotide probe") is an oligonucleotide of approximately 10-50 base pairs, preferably approximately 15-30 base pairs, that specifically hybridizes to the HCRTR2 gene, and that contains a mutation associated with narcolepsy. An allele-specific oligonucleotide probe that is specific for particular mutation in the HCRTR2 gene can be prepared, using standard methods (see Current Protocols in Molecular Biology, supra). To identify mutations in the gene that are associated with narcolepsy, a test sample of DNA is obtained from the individual. PCR can be used to amplify all or a fragment of the HCRTR2 gene, and its flanking sequences. The DNA containing the amplified HCRTR2 gene (or fragment of the gene) is dot-

blotted, using standard methods (see Current Protocols in Molecular Biology, supra), and the blot is contacted with the oligonucleotide probe. The presence of specific hybridization of the probe to the amplified HCRTR2 gene is then detected. Specific hybridization of an allele-specific oligonucleotide probe to DNA from the individual is indicative of a mutation in the HCRTR2 gene, and is therefore indicative of narcolepsy.

Other methods of nucleic acid analysis can be used to detect mutations in the HCRTR2 gene, for the diagnosis of narcolepsy. Representative methods include direct manual sequencing; automated fluorescent sequencing; single-stranded conformation polymorphism assays (SSCA); clamped denaturing gel electrophoresis (CDGE) heteoduplex analysis; chemical mismatch cleavage (CMC); RNase protection assays; use of proteins which recognize nucleotide mismatches, such as *E. coli* mutS protein; allele-specific PCR, and other methods.

PHARMACEUTICAL COMPOSITIONS

The present invention also pertains to pharmaceutical compositions comprising nucleic acids described herein, particularly nucleic acids containing the HCRTR2 gene described herein. For instance, a nucleotide or nucleic acid construct (vector) comprising a nucleotide of the present invention can be formulated with a physiologically acceptable carrier or excipient to prepare a pharmaceutical composition. The carrier and composition can be sterile. The formulation should suit the mode of administration.

Suitable pharmaceutically acceptable carriers include but are not limited to water, salt solutions (e.g., NaCl), saline, buffered saline, alcohols, glycerol, ethanol, gum arabic, vegetable oils, benzyl alcohols, polyethylene glycols, gelatin, carbohydrates such as lactose, amylose or starch, dextrose, magnesium stearate, talc, silicic acid, viscous paraffin, perfume oil, fatty acid esters, hydroxymethylcellulose, polyvinyl pyrolidone, etc., as well as combinations thereof. The pharmaceutical preparations can, if desired, be mixed with auxiliary agents, e.g., lubricants, preservatives, stabilizers, wetting agents, emulsifiers, salts for influencing osmotic

pressure, buffers, coloring, flavoring and/or aromatic substances and the like which do not deleteriously react with the active compounds.

The composition, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents. The composition can be a liquid solution, suspension, emulsion, tablet, pill, capsule, sustained release formulation, or powder. The composition can be formulated as a suppository, with traditional binders and carriers such as triglycerides. Oral formulation can include standard carriers such as pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, polyvinyl pyrollidone, sodium saccharine, cellulose, magnesium carbonate, etc.

Methods of introduction of these compositions include, but are not limited to, intradermal, intramuscular, intraperitoneal, intraocular, intravenous, subcutaneous, oral and intranasal. Other suitable methods of introduction can also include gene therapy (as described below), rechargeable or biodegradable devices, particle acceleration devises ("gene guns") and slow release polymeric devices. The pharmaceutical compositions of this invention can also be administered as part of a combinatorial therapy with other agents.

The composition can be formulated in accordance with the routine procedures as a pharmaceutical composition adapted for administration to human beings. For example, compositions for intravenous administration typically are solutions in sterile isotonic aqueous buffer. Where necessary, the composition may also include a solubilizing agent and a local anesthetic to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the composition is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water, saline or dextrose/water. Where the composition is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

For topical application, nonsprayable forms, viscous to semi-solid or solid forms comprising a carrier compatible with topical application and having a dynamic viscosity preferably greater than water, can be employed. Suitable formulations include but are not limited to solutions, suspensions, emulsions, creams, ointments, powders, enemas, lotions, sols, liniments, salves, aerosols, etc., which are, if desired, sterilized or mixed with auxiliary agents, e.g., preservatives, stabilizers, wetting agents, buffers or salts for influencing osmotic pressure, etc. The agent may be incorporated into a cosmetic formulation. For topical application, also suitable are sprayable aerosol preparations wherein the active ingredient, preferably in combination with a solid or liquid inert carrier material, is packaged in a squeeze bottle or in admixture with a pressurized volatile, normally gaseous propellant, e.g., pressurized air.

Agents described herein can be formulated as neutral or salt forms. Pharmaceutically acceptable salts include those formed with free amino groups such as those derived from hydrochloric, phosphoric, acetic, oxalic, tartaric acids, etc., and those formed with free carboxyl groups such as those derived from sodium, potassium, ammonium, calcium, ferric hydroxides, isopropylamine, triethylamine, 2-ethylamino ethanol, histidine, procaine, etc.

The agents are administered in a therapeutically effective amount. The amount of agents which will be therapeutically effective in the treatment of narcolepsy can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the disease or disorder, and should be decided according to the judgment of a practitioner and each patient's circumstances. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture,

use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use of sale for human administration. The pack or kit can be labeled with information regarding mode of administration, sequence of drug administration (e.g., separately, sequentially or concurrently), or the like. The pack or kit may also include means for reminding the patient to take the therapy. The pack or kit can be a single unit dosage of the combination therapy or it can be a plurality of unit dosages. In particular, the agents can be separated, mixed together in any combination, present in a single vial or tablet. Agents assembled in a blister pack or other dispensing means is preferred. For the purpose of this invention, unit dosage is intended to mean a dosage that is dependent on the individual pharmacodynamics of each agent and administered in FDA approved dosages in standard time courses.

METHODS OF THERAPY

The present invention also pertains to methods of therapy for narcolepsy, utilizing the pharmaceutical compositions comprising nucleic acids, as described herein. The therapy is designed to replace/supplement activity of the hypocretin(orexin) receptor 2 in an individual, such as by administering a nucleic acid comprising the HCRTR2 gene or a derivative or active fragment thereof. In one embodiment of the invention, a nucleic acid of the invention is used in the treatment of narcolepsy. The term, "treatment" as used herein, refers not only to ameliorating symptoms associated with the disease, but also preventing or delaying the onset of the disease, and also lessening the severity or frequency of symptoms of the disease. In this embodiment, a nucleic acid of the invention (e.g., the HCRTR2 gene (SEQ ID NO:1)) can be used, either alone or in a pharmaceutical composition as described above. For example, the HCRTR2 gene, either by itself or included within a vector, can be introduced into cells (either *in vitro* or *in vivo*) such that the cells produce native HCRTR2 receptor. If necessary, cells that have been transformed with the gene or can be introduced (or re-introduced) into an individual affected with the disease. Thus, cells which, in nature, lack native HCRTR2 expression and activity, or have mutant HCRTR2 expression and activity, can be engineered to express

HCRTR2 receptors (or, for example, an active fragment of the HCRTR2 receptor). In a preferred embodiment, nucleic acid comprising the HCRTR2 gene, can be introduced into an expression vector, such as a viral vector, and the vector can be introduced into appropriate cells which lack native HCRTR2 expression in an animal. In such methods, a cell population can be engineered to inducibly or constitutively express active HCRTR2 receptor. Other gene transfer systems, including viral and nonviral transfer systems, can be used. Alternatively, nonviral gene transfer methods, such as calcium phosphate coprecipitation, mechanical techniques (e.g., microinjection); membrane fusion-mediated transfer via liposomes; or direct DNA uptake, can also be used.

The nucleic acids and/or vectors are administered in a therapeutically effective amount (i.e., an amount that is sufficient to treat the disease, such as by ameliorating symptoms associated with the disease, preventing or delaying the onset of the disease, and/or also lessening the severity or frequency of symptoms of the disease). The amount which will be therapeutically effective in the treatment of a particular disorder or condition will depend on the nature of the disorder or condition, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the disease or disorder, and should be decided according to the judgment of a practitioner and each patient's circumstances. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems.

The following Examples are offered for the purpose of illustrating the present invention and are not to be construed to limit the scope of this invention. The teachings of all references cited herein are hereby incorporated herein by reference.

EXAMPLES

EXAMPLE 1 Identification of the Human Narcolepsy Gene

A human BAC library (RPCI11 human male BAC library; see Osoegawa, K. *et al., Genomics 52*:1-8 (1998)) was used. Twenty primers, designed from the mRNA sequence of the HCRTR2 receptor, were employed to identify clones of interest. They are set forth in Table 1.

TABLE 1 Primers Used for Hybridization

#	Name	Primer Sequence	SEQ ID NO:
1	HCRTR2-1-F	TACTACTACTAGGCCACGCG	3
2	HCRTR2-1-R	ACACCAGGAGGAGAAAGCTAC	4
3	HCRTR2-2-F	ATCGCCTGTAAAGACAGCAAAG	5
4	HCRTR2-2-R	AAAGTTACTGAGCCAATGCCTC	6
5	HCRTR2-3-F	GAGAGGAGCTTGCAGCATTG	7
6	HCRTR2-3-R	AGGAATTCCTCGTCGTCATAGT	8
7	HCRTR2-4-F	GAAGAACCACCACATGAGGAC	9
8	HCRTR2-4-R	ATCACTTTGCAAAGGGACTGTC	10
9	HCRTR2-5-F	GTATGCAATCTGTCACCCTTTG	11
10	HCRTR2-5-R	AATGCAGGAGACAATCCAGATG	12
11	HCRTR2-6-F	CAGGCTTAGCCAATAAAACCAC	13
12	HCRTR2-6-R	GATAAGCCAACACCATGAGACA	14
13	HCRTR2-7-F	ACAGATCCCTGGAACATCATCT	15
14	HCRTR2-7-R	CTCGGATCTGCTTTATTTCAGC	16
15	HCRTR2-8-F	CCAATTAGCATCCTCAATGTGC	17
16	HCRTR2-8-R	GTGTGAAAAGGTAAACCAGGCA	18
17	HCRTR2-9-F	CTCAGTGGAAAATTTCGAGAGG	19
18	HCRTR2-9-R	GTTGCTGATTTGAGTGGTCAAG	20
19	HCRTR2-10-F	CTTTCTGAGCAAGTTGTGCTCA	21
20	HCRTR2-10-R	TACCAGTTTTGAAGTGGTCCTG	22

Initial Study with Large Membranes

Four out of 5 membranes having the whole BAC library, containing a total of approximately 160,000 BAC clones representing an approximately 10-fold coverage of the human genome, were used in hybridization studies with these primers. Hybridization was performed with a pool of all 20 primers described in Table 1.

5' End Labeling for Big Membranes

Oligonucleotides were labeled at the 5' end before hybridization, using fresh (less than one month old) [γ^{32}P]ATP (6000 Ci/mmole; 10 µCi/µl). The following protocol is adjusted for 4 membranes in 2 bottles, containing 2 membranes/30 ml of rapid hyb. Each. Briefly, a labeling mixture was made of DNA (8 pmol/µl) (10.0 µl of the primer pool), 10X buffer (12.0 µl), T4 PNK (10 u/µl) (6.0 µl), [γ^{32}P]ATP (30.0 µl, or 600 µCi), and water (62.0 µl) for a final volume of 120 µl. 20 µl of labeling mixture was used per 10 ml rapid hybridization reaction. Incubation of the labeling mixture was for 2 hours at 37°C, followed by transfer to ice, spinning down, and mixing with the rapid hybridization solution. The membranes were prehybridized at 42°C before the labeling mix was added. Sixty µl of the labeling mix was added to each of 2 big bottles containing 2 membranes and 30 ml of rapid hybridization solution.

Hybridization and Washing

The membranes were hybridized at 42°C overnight. After overnight, membranes were washed with 6x SSC, 0.1% SDS at room temperature; washed with 6x SSC, 0.1% SDS at 55°C in a shaking waterbath, repeated until the radioactivity of membranes was lower than 6k using 1x sensitivity; and washed with 6x SSC to remove the SDS. The washed membranes were put in a cassette for overnight exposure at -80°C with a MR single emulsion film. Positive clones were identified and gridded on small membranes.

Study of Positive Clones with Small Membranes

After growing the positively-identified clones on several small membranes (to get several copies of membranes containing the same clones), and washing the membranes, hybridization was performed using pairs of primers, instead of a total pool of primers as before. The total number of hybridizations was ten, using different primers against identical copies of membranes containing all positive clones from the first hybridization. The primer pairs are set forth in Table 2; primer numbers indicate the primers shown in Table 1.

TABLE 2 Primer Pairs Used for Hybridization

Reaction number	Primers Used
1	1 and 2
2	3 and 4
3	5 and 6
4	7 and 8
5	9 and 10
6	11 and 12
7	13 and 14
8	15 and 16
9	17 and 18
10	18 and 19

5' End Labeling for Small Membranes

Oligonucleotides were labeled at the 5' end before hybridization, using fresh [γ^{32}P]ATP (5000 Ci/mmole; 10 μCi/μl). Briefly, a labeling mixture was made of DNA (8 pmol/μl) (1.5 μl), 10X buffer (2.0 μl), T4 PNK (10 u/μl) (1.0 μl), [γ^{32}P]ATP (3.0 μl), and water (12.5 μl) for a final volume of 20 μl. Incubation of the labeling mixture was for 2.5 hours at 37°C, followed by transfer to ice, spinning down, and mixing with the rapid hybridization solution. Membranes were pre-wetted in 6X SSC, rolled in a pipette, and excess liquid drained prior to placing the membrane in the tube. Fifty ml Falcon (polypropylene) tubes were used as container for the hybridization. The membranes were prehybridized at 42°C before 20 μl of labeling mix was added to each tube.

Hybridization and Washing

The membranes were hybridized at 42°C overnight. After overnight, membranes were washed as described above. Four clones which were positive for primers designed using the 5' and 3' end of the mRNA were identified. Clone 403B19 was used to characterize the gene.

Sequencing of Narcolepsy Gene in Clone 403B19

Shotgun sequencing was used to obtain the gene sequence.

Preparation of DNA Samples

BAC DNA was isolated using the Plasmix kit from TALENT-VH Bio Limited. Thirty μg of isolated DNA was fragmented by nebulization: a nebulizer (IPI Medical Products, Inc., no. 4207) was modified by removing the plastic cylinder drip ring, cutting off the outer rim of the cylinder, inverting it and placing it back into the nebulizer; the large hole in the top cover (where the mouth piece was attached) was sealed with a plastic stopper; the small hole was connected to a 1/4 inch length of Tycon tubing (connected to a compressed air source). A DNA sample was prepared containing 30 μg DNA, 10 X TM buffer (200 μl), sterile glycerol (1 ml), and sterile dd water (q.s.) for a total volume of 2 ml. The DNA sample was nebulized in an ice-water bath for 2 minutes and 40 seconds (pressure bar reading 0.5). The sample was then briefly centrifuged at 2500 rpm to collect the DNA; the entire unit was placed in the rotor bucked of a table top centrifuge (Beckman GPR tabletop centrifuge) fitted with pieces of Styrofoam to cushion the nebulizer. The sample was then distributed into four 1.5 ml microcentrifuge tubes and ethanol precipitated. The Dried DNA pellet was resuspended in 35 μl of 1X TM buffer prior to proceeding with fragment end-repair.

Fragment End Repair, Size Selection and Phosphorylation

The DNA was resuspended in 27 μl of 1X TM buffer. The following materials were added: 10 X kinase buffer (5 μl), 10 mM rATP (5 μl), 0.25 mM dNTPs (7 μl), T4 polynucleotide kinase (1 μl (3 U/μl)), Klenow DNA polymerase (2 μl (5 U/μl)), T4 DNA polymerase (1 μl (3 U/μl)), for a total volume of 48 μl. The mixture was incubated at 37°C for 30 minutes, and then 5 μl of agarose gel loading dye was added. The mixture was then applied to separate wells of a 1% low melting temperature agarose gel and electrophoresed for 30-60 minutes at 100-120 mA. The DNA was then eluted from each sample lane, extracted from the agarose

using Ultrafree-DA columns (Millipore) and then cleaned with Microcon-100 columns (Amicon), precipitated in ethanol, and resuspended in 10 μl of 10:0.1 TE buffer.

Ligation

EcoRV-linearized, CIAP-dephosphorylated Bluescript vector was used as a cloning vector. The following reagents were combined in a microcentrifuge tube, and incubated overnight at 4°C: DNA fragments (100-1000 ng), cloning vector (2 μl (10 ng/μl)), 10X ligation buffer (1 μl), T4 DNA ligase (NEB 202L) (1 μl (400 U/μl)), sterile dd water (q.s.), for a total of 10 μl.

Transformation of Ligated Products

The ligation products were diluted 1:5 with dd water and used to transform electrocompetent TOP 10F cells (Invitrogen) using GenePulser II (Biorad; voltage, 2.5 W, resistance 100 ohm). Transformants were plated on LB plates with 50 μl of 4% X-GAL and 50 μl of 4% IPTG, and ampicillin. Transformants were grown overnight at 37°C, white colonies were picked, grown in a culture of 3 ml LB liquid media plus 200 μg/μl ampicillin for 16-20 hours with shaking. DNA was isolated from the liquid cultures using Autogen 740 Automatic Plasmid Isolation System.

Cycle Sequencing of Isolated Plasmid DNA

Isolated plasmids were then sequenced using the M13 primers: M13-forward (SEQ ID NO:23) TGTAAAACGACGGCCAG; and M13-reverse (SEQ ID NO:24) CAGGAAACAGCTATGAC. For the sequencing reaction, 2.5 μl plasmid template was mixed with 4 μl Big Dye Ready reaction mix (ABI), 1 μl of 8 pM M13 primer, and 2.5 μl dd water. For cycle sequencing, 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds, and 60 °C for 4 minutes were performed, followed by holding at 4°C. The cycle sequencing reaction products were cleaned by spinning through Sephadex G-50 columns. The eluted cycle sequencing products were then dissolved in 3 μl formamide/dye and 1.5 μl of sample was loaded on ABI 377 automated sequencers. The data was analyzed using Phred and Phrap (Ewing, B. *et al., Genome Res. 8*:175-

185 (1998); Ewing, B. and Green, P., *Genome Res. 8*:186-194 (1998)), and viewed in Consed viewer (Gordon, D. *et al., Genome Res. 8(3)*:195-202 (1998)).

Analysis of Gene Structure

The *hcrtr-2* gene maps to chromosome 6p11-q11. A total of 168,575 base pairs of contiguous sequence was generated for 403B19 which contained all of the *hcrtr-2* gene. Comparison of the cDNA sequence of *hcrtr-2* (Accession number GI:6006037) and the genomic sequences generated allowed deduction of the intron/exon organization of the gene. The gene contains 7 exons which cover 108,439 bp. The first 10 Gs in the mRNA sequence for *hcrtr-2* were not found in the genomic sequence. It is likely that these Gs were an artifact.

The splice junctions of the *hcrtr-2* gene are set forth in Table 3, and the intron sizes are set forth in Table 4. Exon sequences are represented in uppercase and introns in lowercase. All splice sites conform to the consensus GT-AG rule. SEQ ID NOs are given in the column immediately following each site.

Table 3 Splice Junctions of *hcrtr-2*

	Splice Donor Site	SEQ ID	Splice Acceptor Site	SEQ ID
Hcrtr-2 exon1-2	TCCTGGgtgagt	25	aattagTTTGTG	26
Hcrtr-2 exon2-3	CTACAGgtaatt	27	ctctagACCGTG	28
Hcrtr-2 exon3-4	GGGGTGgtaagt	29	tcctagGTGAAA	30
Hcrtr-2 exon4-5	CGACAGgtatat	31	tttcagATCCCT	32
Hcrtr-2 exon5-6	AAAGAGgtaaaa	33	ctgcagAGTATT	34
Hcrtr-2 exon6-7	TCAGTGgtgagt	35	tgccagGAAAAT	36

Table 4 Intron Sizes of *hcrtr-2*

Intron	Nucleotides
Intron 1	73,848
Intron 2	6,322
Intron 3	8,327
Intron 4	13,618
Intron 5	2,730
Intron 6	1,779

The exons do not clearly respect the domain structure of this seven membrane domain G protein linked receptor. Five of the transmembrane regions are by themselves within one exon, two of the transmebrane segments are broken up by introns, and two transmembrane segments fall within the same exon. A survey done one year ago on mammalian G-protein coupled receptors (GPCRs) sequences in GenBank revealed that over 90% of GPCRs genes were intronless in their open reading frame (ORF) (Gentles, A.J. and Karlin, S., *Trends Genet. 15*:47-49 (1999)). Comparison of the intron/exon boundaries of *hcrtr-2* and the genes coding for their most related GPCRs based on sequence similarity showed that the location of the intron/exons boundaries with respect to the transmembrane domains is only partially conserved among the receptors (Sakurai, T. *et al.*, *Cell 92*:573-585 (1998)).

Computer analysis of sequence data

Analysis of the genomic sequence of *hcrtr-2* using the program RepeatMasker (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker) showed that the sequence containing the *hcrtr-2* genomic sequence is 38.27% repeat sequences and the GC content is 35.3%.

The sequences of the genes were analyzed using the program GeneMiner (Óskarsson and Pálsson, unpublished), which combines the results of 5 exon prediction programs; FGENE (Solovyev, V. and Salamov, A., *Ismb 5*:294-302 (1997)), Genscan (Burge, C. and Karlin, S., *J. Mol. Biol. 268*:78-94 (1997)),

HMMgene (Krogh, A., *Ismb 5*:179-186 (1997)), MZEF (Zhang, M.Q., *Proc. Natl. Acad. Sci. USA 94*:565-8 (1997)) and Xpound (Thomas, A. and Skolnick, M.H., *IMA J. Math Appl. Med. Biol. 11*:149-160 (1994)). For *hcrtr-2,* 3 out of 5 programs predicted the 3' end of exon 1, only one program predicted the 7th exon and for the internal exons, there were at least two programs that predicted each of them exactly or in part.

The promoter sequences of the genes have not yet been characterized. The Promoter Prediction by Neural Network (http://www.fruitfly.org/seq_tools/promoter.html) predicted promoters that are at least 140 bp upstream of the 5'UTR of *hcrtr-2*, indicating that either a part of the 5'UTR is missing in the published mRNA sequence or the real promoters are not detected by the program.

Analysis of Population for Polymorphisms

Each exon and its flanking intronic sequences of the *hcrtr-2* gene was analyzed in nucleic acid samples from 47 patients and 75 control individuals. The patient population consisted of patients of Icelandic and US origin. The control population consisted of Icelandic controls, CEPH (Centre d'Etude du Polymorphisme Humain) individuals from Utah and France, and US samples of various ethnic origins. The African-American/Caucasian ratios were similar between patients and controls. All narcoleptic subjects complained of excessive daytime sleepiness (EDS). Approximately 66% of the patients had cataplexy, 24% did not and 10% did not have attainable records of cataplexy status. Narcoleptic subjects without cataplexy had Multiple Sleep Latency Tests showing mean sleep latencies of less than 10 minutes and REM sleep in at least 2 naps. Subjects did not take any drugs affecting sleep for at least 10 days before their sleep studies.

To analyze the nucleic acids, DNA from patient and control blood samples were isolated by the method of Kunkel (Kunkel, L.M. *et al., Proc. Natl. Acad. Sci. USA 74*:1245-9 (1977)). Briefly, white blood cells were lysed in a sucrose lysis buffer, and proteinase K treated; the DNA was then extracted using phenol-chloroform/isoamylalcohol and then ethanol precipitated. Patient samples that were

received in the form of nuclei pelleted through sucrose buffer were resusupended in lysis buffer (100 mM NaCl2; 10 mM TrisHCl, pH 8; 25 mM EDTA pH 8; 0.5% sodium dodecyl sulfate; 0.1 mg/ml proteinase K) at 55°C for 4-6 hours followed by classical phenol-chloroform extraction and ethanol precipitation (Sambrook, J. *et al.*, *Molecular Cloning, A Laboratory Manual* (1989)). Samples were incubated at 55°C after isolation for the inactivation of DNAse to prevent the degradation of DNA. Concentration of the isolated DNA was determined by spectrophotometric analysis at 260 nm (Sambrook *et al.*,using GeneQuant (PharmaciaBiotech), and samples diluted with sterile distilled water to a 20 ng/µl working solution.

Primers were designed from intronic sequences flanking the exons of the hypocretin receptor-2 (*hcrtr-2*), using either primer design programs available at primer3 at the Whitehead Institute (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi) or primers for the worldwide web (http://williamstone.com/primers/javascript/). The primers are shown in Table 5.

Table 5 Primers Used to Amplify Nucleic Acid Fragments for Analysis of *hcrtr-2* Gene

EX-ON	#	Primer Sequence	Sense/ Antisense	External/ Nested	SEQ ID.
1	1	TTTCTTCAGCTTCAGCTCTCCCTCAGC	S	E	37
1	2	TTCAGCTCCGAAGCAGATGACCAGTTG	A	E	38
1	3	TTCAGCTTCAGCTCTCCCTCAGCGAGG	S	N	39
1	4	CGAAGCAGATGACCAGTTGCGACAAGG	A	N	40
1	5	CTTTCCCACCGCAAATCACCAGTGCTC	S	E	41
1	6	ATTTTATTAGAAAACCCCATCCGAGAG	A	E	42
1	7	TTCCCACCGCAAATCACCAGTGCTC	S	N	43
1	8	TATTAGAAAACCCCATCCGAGAGCAG	A	N	44
2	9	GCATGTACTTAGCATTCACACAGATTG	S	E	45
2	10	TCTAATGATGATTTGGCAGTTCATTGC	A	E	46
2	11	TAGCATTCACACAGATTGACAGATTCA	S	N	47
2	12	CAGTTTGTCAATGCCTTAGGCAAATAT	A	N	48
3	13	TTTGGCAGCTTTGAATTTGCTTATATG	S	E	49
3	14	GCTCTTGCAAAACTGTATTCACAAATG	A	E	50
3	15	CAGCTTTGAATTTGCTTATATGTTGTG	S	N	51
3	16	TTGCAAAACTGTATTCACAAATGTCAA	A	N	52
4	17	TCCCCTTTGCATACATAATATGACAATG	S	E	53
4	18	AAAAAGCACAGACAAAATATTTGGAAGG	A	E	54
4	19	ATGCACTTTGAAGAAAAGCATTGACATG	S	N	55
4	20	AAGCACAGACAAAATATTTGGAAGGAAT	A	N	56
5	21	CTCAGGCGTCTGGAAGCCTTTCCTTAC	S	E	57
5	22	TTAAAGGCTGTTCGCCTTACCTGCTGG	A	E	58
5	23	GGCGTCTGGAAGCCTTTCCTTACTGTG	S	N	59
5	24	CTGAGTCATCTGGCCTGACAAGGTATC	A	N	60
6	25	GGGTCAGAAACCAATCTGTGGTCAATTC	S	E	61
6	26	AGTTGAAGAGTGTTCATTGATTCCTCATCC	A	E	62
6	27	AGAAACCAATCTGTGGTCAATTCCTGCAAC	S	N	63

EX-ON	#	Primer Sequence	Sense/ Antisense	External/ Nested	SEQ ID.
6	28	TGAAGAGTGTTCATTGATTCCTCATCCTTG	A	N	64
7	29	GAGTCTACCAAGCTTCCAATAAACTCA	S	E	65
7	30	GGATAGTTTTACTCAGGTATCCTTGTCA	A	E	66
7	31	CAAATCAGCAACTTTGATAACATAT	S	N	67
7	32	GTATCCTTGTCATATGAATAAATATTCTAC	A	N	68
7	33	CACTCAAATCAGCAACTTTGATAAC	S	E	69
7	34	GTGAGAGATTAAAATAACAAGGGAT	A	E	70
7	35	CAAATCAGCAACTTTGATAACATAT	S	N	71
7	36	TGTTTAAACATTTAATTGACACACA	A	N	72
7	37	TTCATATGACAAGGATACCTGAGTAAA	S	E	73
7	38	GTGAAATAGCCTGAAATAAGCTCAA	A	E	74

PCR reactions were done in 20 µl reactions using 40 ng genomic DNA, 0.2 mM solution of the four dNTPs, 0.35 µM of each primer (TAGCopenhagen), 2.5 mM MgCl2 (Perkin Elmer), 1x PCR Buffer (Perkin Elmer) and 0.5 U Amplitaq gold (Perkin Elmer). The primers were used to amplify the fragments by PCR cycling at 95°C for 12 min and subsequently 30 cycles of 95°C for 30 sec, 55-62°C for 30 sec and 72°C for 1 min. The PCR products were prepared for cycle sequencing by incubation with Shrimp alkaline phosphatase (Amersham) and exonuclease I (Amersham) at 37°C for 15 min. After the inactivation of the enzymes the products were subject to cycle sequencing using BigDye Ready Reaction mix (Perkin Elmer) and subsequently run on ABI Prism 377 Automated DNA sequencers. The raw data were basecalled and sequences assembled using the Phred and Phrap software, respectively (Ewing, B. *et al., Genome Res. 8*:175-185 (1998); Ewing, B. and Green, P., *Genome Res. 8*:186-194 (1998)). The Consed viewer was used to analyze the sequences (Gordon, D. *et al., Genome Res. 8(3)*:195-202 (1998)). Expansion of a T-stretch in the 3' untranslated region (UTR) of exon 7 of *hcrtr-2* was investigated by amplifying a fragment containing the stretch with a fluorescently labelled primer

pair using the conditions described above. The PCR product was dissolved in formamide/dye solution and run on ABI Prism 377 Automated DNA sequencers as described above. Allele calling was done using TrueAllele and editing was done using DeCODE-GT (Palsson, B. *et al., Genome Res. 9*:1002-1012 (1999)).

A total of nine single nucleotide polymorphisms were identified, 7 in exons and 2 in an intronic sequence near an exon. The polymporphisms are shown in Table 6. The base number is according to the mRNA sequence (Accession number GI:6006037). For those polymorphisms marked with an asterisk (*), the polymorphism is located 5' of the corresponding exons; the numbers indicate the distance into the introns.

Table 6 Single Nucleotide Polymorphisms in *hcrtr-2*

Location	cDNA base #	Nucleic Acid Change
Exon 1	352	C-T
Exon 1	355	C-A
Intron1	-26*	C-A
Exon 5	1,170	G-A
Exon 5	1,177	C-A
Exon 5	1,201	G-A
Exon 5	1,246	G-A
Exon 5	1,266	G-A
Intron 6	-87*	G-A

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the spirit and scope of the invention as defined by the appended claims.

CLAIMS

What is claimed is:

1. Isolated nucleic acid molecule comprising the nucleic acid having SEQ ID NO:1.

2. A DNA construct comprising the isolated nucleic acid molecule of Claim 1 operatively linked to a regulatory sequence.

3. A recombinant host cell comprising the isolated nucleic acid molecule of Claim 1 operatively linked to a regulatory sequence.

4. A pharmaceutical composition comprising a nucleic acid comprising the isolated nucleic acid molecule of Claim 1.

5. Isolated nucleic acid molecule comprising the nucleic acid having SEQ ID NO:1 with one or more of the nucleic acid changes shown in Table 6.

6. A method of diagnosing narcolepsy in an individual, comprising detecting a mutation in the gene encoding hypocretin (orexin) receptor 2, wherein the presence of the mutation in the gene is indicative of narcolepsy.

7. A method of treating narcolepsy in an individual, comprising administering to the individual an isolated nucleic acid of Claim 1 in a therapeutically effective amount.

```
LOCUS       __________          168,575 bp    DNA      PRI     20-OCT-1999
DEFINITION  Human hypocretin (orexin) receptor 2 (HCRTR2) gene, complete cds.
ACCESSION   __________
NID         __________
VERSION     __________
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1-168,575)
AUTHORS     __________
  TITLE     Direct Submission
  JOURNAL       Submitted (__________) deCode Genetics, Inc., Lynghals 1,
                IS-110   Reykjavik, Iceland.
FEATURES        Location/Qualifiers
     source     1.. 168,575
                /organism="Homo sapiens"
                /db_xref="taxon : 9606"
                /chromosome="6"
                /map="6p11-q11"
                /clone="BAC 403B19"
     gene       1..129,305
                /partial
                /gene="HCRTR2"
                /note="OX2R"
                /db_xref="LocusID:3062"
                /db_xref="MIM:602393"
     exon       20,867..21,403
                /gene="HCRTR2"
                /number=2
     CDS        join(21,181..21,403, 95,252..95,430, 101,753..101,996, 110,324..110,439,
                124,058..124,278, 127,009..127,130, 128,910..129,139)
                /gene="HCRTR2"
                /note="HCRTR2 exons defined by comparison to mRNA sequence (NM_001526)"
                        /product="HCRTR2/orexin 2 receptor"
                /db_xref="LocusID:3062"
                /db_xref="MIM:602393"
                /protein_id="NP_001517.1"
                /db_xref="PID:g4557639"
                /db_xref="GI:4557639"
                /translation="MSGTKLEDSPPCRNWSSASELNETQEPFLNPTDYDDEEFLRYLW
                REYLHPKEYEWVLIAGYIIVFVVALIGNVLVCVAVWKNHHMRTVTNYFIVNLSLADVL
                VTITCLPATLVVDITETWFFGQSLCKVIPYLQTVSVSVSVLTLSCIALDRWYAICHPL
                MFKSTAKRARNSIVIIWIVSCIIMIPQAIVMECSTVFPGLANKTTLFTVCDERWGGEI
                YPKMYHICFFLVTYMAPLCLMVLAYLQIFRKLWCRQIPGTSSVVQRKWKPLQPVSQPR
                GPGQPTKSRMSAVAAEIKQIRARRKTARMLMVVLLVFAICYLPISILNVLKRVFGMFA
                HTEDRETVYAWFTFSHWLVYANSAANPIIYNFLSGKFREEFKAAFSCCCLGVHHRQED
                RLTRGRTSTESRKSLTTQISNFDNISKLSEQVVLTSISTLPAANGAGPLQNW"
     exon       95,252..95,430
                /gene="HCRTR2"
                /number=3
     exon       101,753..101,996
                /gene="HCRTR2"
                /number=4
```

FIG. 1A

```
     exon            110,324..110,439
                     /gene="HCRTR2"
                     /number=5
     exon            124,058..124,278
                     /gene="HCRTR2"
                     /number=6
     exon            127,009..127,130
                     /gene="HCRTR2"
                     /number=7
     exon            128,910..129,305
                     /gene="HCRTR2"
                     /number=8
```

BASE COUNT 55,308 a 29,672 c 29,838 g 53,757 t

```
CGACTTGATTTTATTTTTTGCATATGGATATCCAGTTTTCACAGCACTGCTTGTTACCCT
CAGCAAAGAACAGTTGTCTGTAAATTCATGGGTTTATGTCTAGGCTCTCTGTTCTGTTCT
ATTGGTCAACATATGGTCATATATCACTTAACTGCAGGGAAGGGATACATTCTGAGAAAT
GCATTATTACATGATTTCATCATTGTGCAAACACTATAGAGTGTAGTTACAGAAACCTAG
TATCTCTAGCTGTGTTCTTATGATTCAAATTTGCTTTGGTCATTTGAGATCCATACTGGT
GGAGTCTAATTATTCAAAACTAGGGAAAACAGACAAACAGAAAAAACTAAGACCAAGTTA
GCAGAAGAAAGACAATAACAAAGGTTAGATCAAAAATAAATAATATAGAGAATGAAAAAA
TTAGAAAAAGTGGACAAAACTACAATGTACTTTTTTGAAAAGACAAACAAAATTAACAAA
CCCTTACCTTGACTAAAAAAAGAGACTCAAATAAATAAAATTGGAAATGAGACAGGAGAC
ATTACAATTGATGTTAACAAAAAGATCATAAGGTACTATTATGAACAACTATACACCAAT
AAATTGGACAACCTAGAAAAAAAATGGATAAATTCCTAGAAATACACAGTCTATCAAACT
GAAACAAGAAGAAATAGAAAGCCTGAACATACCAGTAACAACCAAGGAGACTGAGTAAAT
AATCAAAAACCTCCCAAGAAGAAGAGTCTAGGACCAGAAGTCTTCACAAATGAATTCTAC
CAAACATTTAAAGTATTAATGCCAATCATTCATTCTTATACTCTTCCAAAAAGAAAGAGG
GAATATTTTCAAACTCATTTTATGAGGCCAGCATTATTCTGATACCAAAACTACGCAAAA
ATACTACAAGAACATAAAAACTACAAATGTGGGAATTATCATGTATACATATGCAAAAAT
CCTCAGTAAAATCCTAGCAAACTAAATTCAACAGTACATTAAAAAGATCATATAGCATGA
CCAGTGAAATTTCTCCTTAGGACGCAAGGATAAGTCAACATATAAAATTGAATGTGATAT
ACCACTTTAACAAAATGAAGGATAAAAATCATATGATCATCTGAATAGATGCAGAAAAAG
CATATAACAAACTTTGACGTTGTTGAGAAATTGAAAGCTTTTCTCTAAGATCAAGAACAA
AGCAAGGATGCCCATTCTTGCTTCTATTCAGCATAGTGCTTGAAGTCCTAGTCTGGACAA
TTGGGCAAAAATAAATAAATAAATAAATAGATAAATAAATAAATAAATATAAATAAATAA
ATAAAATCCACCAAATTGGAAAGGAAGAAGTGAAATTACCTCTGTTTGTAGATGAGCTGA
TCTCATGTGTAGACAACCTTAAAGATTCCACAAAAACAAACAAACACACAAACAAACAAA
ATAGCTAGAGCAAAGAAATGAATTCAGTACAGTTGCAGAATGCAAAATCAGTATACAAAA
AGTACTTGTAATTCTATATAATAGCAACAAACTATTTCATAAGGAAATTAAGGAAACAAT
CCCCATTACAATAGCATCATAAATAATAAATCTTAAGAACAAATTTAACCAAGGAGGTGA
AAGACTTGTGTACTGAAAACTATAAAATGCTGATAAAAAAATTAAAGAAGATACAATAAA
TGGAAGATATTCCATGCCATGGTTTGGAAGAATTAATATTGCTAAATGTACATACTACCC
AAAACAACCTGTAGAGTCAATGCAATCCCTATCAAAATACCAATATTTTTTTTTTTACAG
AAATGAAAACACAATCTTAACACTATTTAAACCAATTAAACAAACCTATGATTTCAATTT
GGTCAAATGTGTTAGAATGGATTTCCTTTTATTGTTTTGAACTTGTCTCTTCCAAATTTC
AAAGCCTGGTTCCTAATTTTTACTTGAAATACCAAATAACAAACCCACTTAATGAGCTCT
GAGCCAGTTTTAGTAGCCAAACTTGTATTTAAATAGTGTGTTACATATTTGCACAAAAAG
CCAACGGAGTCTAAATCAACACTAATTCACATCATTACTAGCAATCTAAAACATCAGATG
ATAATTTTGCTGTTGTCTTTCAGGCAAGATATTCAACCATTGGTATTAAATGTTTTATAT
GAATGTGCGGTGTTTTATTTCAGAAACACTTCTCTGAATTCCCAAGGCCTAAGAGCTATT
CATCATAGAGGTTTGTGGAGGCGGTAGTTAGACATTTTCTACATGCATAATGTTAATTCA
TTCAAACATTATAGAAAAAAAGTTTGTAAAGAAGTTAATTTTCAAGGTGACAAAAAAATC
AGATTGAATCATGTTTATTTTATTTCAATTTAAACTCGTTGGCTATCTTAGGAAATTCAC
ATTGTTTTTGAAGAATATATGAACAAAGTTTGATTCATCTTATCTATATAAGCATGAGAG
```

FIG.1B

AAATACCTAGATGAGAGTGAAAAATGACTAATTTTGTGACCATTGTTATATCATAGATTA
ACTTGTTCTCTTCTACTTCTAAGCTGTGTGATCTTGAAAAGTCATCTAAACTTCAGGTAC
CATCCTCACTTGCAAAATGAGGGGAAAAACCCCAGCACCTTTAATATGGTGTTATGTGGA
TGAAATAAGTTAATATATATTAAGTGCTTAGGTTTCATGCACTTTCTATATAGTATTAAT
AATATTATTGTTACATTATTATAGTTACATTATTATTTTTATTAATATTATTGGAACATG
AATGGAATTGTTGTGGCTCATTTTAAAGATGCTGCAATGGAGACCAAGAGAAATTAAGTA
TAATAATCCTAGTTAAGGTACAGCCATTTCTAATTACATTTTCCAACTGCTGCTTTTACT
CCTAGCACTCACACCAATTCTTCTCATAATCTAATAAATACTGAAATTTAAAACTTATAA
AGAACACATAATAATCTTATTTAATTATCACAACAATTTCTGTGGAGTTACTATTAATCC
AGAGATGAAGAATCTAAAACTCTAAATTATCAAGCAACTATTCCAGCTTTAAACAACAGT
AAAACTGGAATTAAAACTAGAGTTTCTTTATGAGGCCAGTATTACTCTTATACTAAAGCA
AGACAACTGTCTCTCTCTCACTCTCTCTTTCTCTCTCTCTGAGACACACACACACACACA
CAAGCACACACACACACACACACCAGATCAATATAACTTATGGATGTTGATGCAAAAATTTT
CAAAAAATTAATAGCAAATCGAATCCAGCAGTATTTTAAAGGACTATACACCATGGACAA
ATGGGGTTTATTCCTGGGATATAAAGTTGGCTAAACTTAATGAAAATCAACCAGTGCAAT
AAATAATAGTAATTAAAAAAACATAATTATCTCACTAGATGTACAAAAAATGACAAGATC
CAACATGTTTTCAATATAAAAGCATTCCACAGACTAGGAATAGAAGGGAACTTCCTCAAC
TTTACAAAGAACATCTACAACGAAACCACAGCTAACATCATATTTAATGGTGGAAGACTG
AAATCTCTGAATATTTTCCCCTAAGATCAGAAAAAGACAAAGAGGTCTACTAATTCTATT
CAACATTGGAAAGATAGTTCTAGTCAATTAATCATTAAAAAAAGGCATTTAGATTAGAAA
GGAAGTAAAATTACCTCTGGCAGATGACATAATCTTATACATGGAGAACTCCTAGAGATT
ACATACACACTCACAACTACCAGAGTTAATAAATGGGTTCTACAAAGTTGCAGGATACAA
TATCAATTCTCAAAAAACACTTGTATTTCTATACACTAGCAAATAACTCTGAAAATGAAA
ATAACAAAACAATTCCACTTGCAAGAGCATCAAAAAAGCATGAAAATCTTAGGAAGGAAT
TTACCAAGAATGTACAAGTTTTATGTACTGAAAAATAAAAAATGTCATTAAAAATAGTTA
AAGAGAATCTAAGCACATTGCAGTTTTCTGACTCCAGGCCCGGGCTCTTGGATGGCATCT
CTGGATCCACTCAGGACCAGGGAGAACTTGTTGCCCTGAAGGGAAGGACACAAGTCTGAC
TGGCTTTACCACCTGCTGATTGTAGAAGCCTAGGGCCTTCAGGGAACACAGGTGGTAGCC
AGATAGCAGTTACCATGGGCATTAGGCATGACCCAGTGCTATGTTGGCTTCTAGTTTGAC
CCAGCACAGCCCAAGGGTGGTAACCACATGGGTGCTTGTGTCACCCCTCCTTAAGTTCCA
GGCAGCCAGCAAAGAGAGAGTGACTCTGTTTGGGAGAAAGTAAGGGAAGAGAATAAAAGT
CTCTGTTGGTAATACAAGGAATTCTTCCAGATCTTATCCAAGACCTCTATGAATCTGCAA
CAGCCAAAGCATTATTAGTTTTCAGGTTTCCCCAGTGCAGATATGACTGCAATGATCAAA
AACTTAGATTATAACACTCAAGTCCCATTTGATACCTGAAAAGCTTTCCAAGAAAGATAG
GCACAAACAAGTTTGGACTGGGAGGACTACAATAAATACCTAACTTCTCAATGCCCAGAA
ACTGATGAACATCCACAAGCTTTAAGACCATCCAGGAGGTGGCTGGCAAGATGGCTGAAT
AGGAACAGCTCTGGTCTGCAGCTCCCAGTGAGATCAGTGCAGAAGGTGGTGGTTTCTGCA
TTTCCAACTGAGGTACCCAGCTCCTCTCATTGGGACTGGTTAGATAGTGAGTGCAGCCCA
CAGAGGGTGAGCCAAAGCTGGATGGGGTGTCACCTCACTGGGGAAGCACAAGGGGATGGG
GAACTCCCTCCCCTAGCCAACGGAATCTGTGAGGGACTGCCATGAGGGATGGTGCATTCT
GGTCCAGATACTATGCTTTTCCCATGTTCTTCACAACCCTCAGGCCAGGAGATTCCCTCG
GGTGCCTACACCACCAGGGCCTTGGGTTTCAAGTACAAAACTGGGTGGATCTTTGGGCAG
GCACCGAGCTAGCTGCAGGAGTTATTTTTCATACCCCAGTGGTGCCTGGAATGCCAGTGA
GACAGAACCATTCACTCTCCTGGAAAGGGAGCTGAAGCCAGGGAACCCAGTGGTCTAGCT
CGGTGGATCCCACTCCCATGGAGGCCAGTAAGCTAAGATCCACTGGCTTGAAATTCTCAC
TGCCAGTGCAGCAGTCTGAAGTCAACCTGGGATGCTTGAGCTTGGTGGAGAGAGGGACGT
CCACCATTACTGAGGTTTGAGTAAGCAGTTTTCCCCTCACAGTGTAAACAAAGCCACTGG
GAAGTTAAAGTAGGTGGAGCCCACGACAGTTCGGCAAAGCCACTATAGCCAGAATGCCTC
TCTAGATTCCTCCTGTCTGGGCAGGGCTTCTCTGAAAGAAAGGCAGCAGCTGCGGTCAGG
AGCTTATAGATCAAACTCCCATCTCCCTGGGACAGGGCACCTGGGGAAAGGGGCAGCTGT
GGGTGCAGCTTCAGCAGACTTAAATATTGCCGCAAGCTGACTCTGAAGACAGCAGGGGAT
CTCCCAGCACAGCGCTCGAGCTCTGCTAAGGGGCAGACTGCCTCCTCAAGTGGGTCTCTG
ACCCCTGTGTCTCCAGACTGGGAGACACCGCACAGCAAGGGTCGACAGACACCTCATACA
GGAGAGCTCCGGCTGGTATCTGGTGGGTGCCCCTCTGGGACAAAGCTTCGAGAGGAAGGA
GCAGGCAGCAATCTTTGCAGTACTGTAGCCTCTACTGGTGATACCCAGGCAAATAGGGTC
TGACGTTGACCTCCAGCAAACTCCAGCAGACCTTCAGCAGACGGGCCTGAGTGTAAGAAG
GAAAATTAACAAACAGAAAGGAATAGCATCAACATCAAAAAAACAAAAACAAAAACAAAA

FIG. 1C

ACAAAAACAAAAACAGCACATCCGCACAAAAACCCCATCTGAAGGTCACCAACACCAAAT
ACCAAAGGTAGATAAATCCACAAAGATGGGGAAAAACCAGCACAAAAAAGCTGAAAATTC
CAAAAAACAGAATACCTCTTCTCCTCCAAAGGATCACAATTCCTCACCAGCAAGGGGACA
AAACTGGACAGAGAATGAGTTTGATGAATTGACAGAAGTAGGCTTGAAAAGGTGGGTAAT
AAACTCCTCTGAGCTAAAGGAGCATGTTCTAACCCAATGCAAGGAAGCTAAGAACCTTGA
AAAATGGTTAGAGTAATTGCTAACTAGAATAACCAGTTTAGAGAAGAGCATAAATGACCT
GATGGAGCTGAAAACTATAGCACAAGAACTTCGTGCAGCATACACAGGTATCAATATCCA
AATCGATCAAGCAAAGAAAAGAATATCAGAGATTGAAGATCAACTTAATGAAATAAAGTG
TGAAGACCAGATTAGAGAAAAAAGAATAAAAAGGAATGAACAAAGTCTCCAAGAAATATG
GGAATATGTGAAAAGACTAAACCTACATTTGATTAGTGTACCTGAAAGTGACGGGGAGAA
AGGAATCAAGTTGGAAAACATTCTTCAGGATATTATCCAGGAGAACATCCACAACCTAGC
AAGACAGGCCAACATTTAAATTCAGGAAATACAGAGTACATCACAAAGATACTCCTCGAG
AAAAACAACCCCAAGACACATAATTGTCAGATGCACCAAGGTTGAAATACAGGAAAAAAG
TTAAGGGCAGCCAGAGAGAAAGGTCGGGTTACCCACAAAGGGAAGCCCATCAGACTAACA
GTGGATCTCCCTGCAGAAACCCTACAAGCCAGAAGAGAGTGAAGGCCAATATTCAACATG
CTTTAAGAAAAGAATTTTCAACCCACAATTTCATATCCAGCCAAACTATGCTTCATAGTG
AAGGAGAAATAAAATCCTTTACAGACAAGCAAATGCTGAGAAATTTTGTCACCACCAGGC
CTGCCTTACAAGAGCTCCCGAAGGAAGCACTAAATATGAAAAGGAAAAACCAGTATCAGC
CACTGCAAAAACATATGAAATTGTAAAGACCATCAACACTATGAAGAAACTGCATCAACT
AATGGGCAAAATAACCAGCTAGCATTATAATGACAGGATCAAATTCACACATAACGATAT
TAACCTTAAATGTAATAGGCTAACTGCCCCAATTAAGAGACACAGACTGGCAAATTGGAT
AGAGAGTCAAGACCCAACAGTGTGCTGTATTCAGGAGTCCAATTCATGTGCAAAGATACA
TATAGGCTCGAAATAAAGGGATGGAGGAATATTTACTAAGCAAATGGAAAGCAAAATAAA
GCGGAGGTTGCAATCCTAGTCTCTGATAAAATAGACTTCAAACCAACAAAGATCAAAAGA
GACAACAAAGGGCATTACATAATGATAAAGGGATCAATGCAACAAGAACAGCTAGCTATC
CTAAATATATATGCACCCAATTCAGGAGCACACAAATTCATCAAGCAAGTTCTTAGAGAC
CTATAGAGACTTAGACTCCCACGTAATAATAGTGGGAGACTTTAACACCCCACTGTCAAT
ATTAAACAGATCAATGAGACAGAAAATTAACAAGTACATTCAGGACTTGAACTCAGCTCT
GGACCAAGCAGGCCTAATAGACATCTATAGGACTCTCCACCCCAAATAAATAGAATATAC
ATTATTCTCAGCACCACATTGCACTTATTCTAAAATTGACCACATCATTGGAAGTAAAAG
ACTCCTCAGCAAATGCCAAAGAACTAAAATCATAACAAACAGTCTCTCAGACCACAGTGC
AATCAAATAAGAGCTCTGGAATAAGAAACTCACTCAAAACCGCACAACTACATGGAAACT
GAACAACCTGCTGCTGAATGACTACTGGGTAAATAATGAAATTAAGGCAGAAATAAATAA
GTTACTTGAAACCAATGAGAACAAAGACACAACATACCAGAATCTCTGGGACACAGCTAA
AGTAGTGTTTGGAGGGAAATTCATAGCACTAAATGCCCACACGAGAAAGTGGGAAAGATC
TAAAATCAACACCCTAACATCACAATGAAAAGAACTAGAGAAGCAAAGGCAAACAAATTC
AAAAGCTAGCAGAAGACAAGAAATAACTAAGATGAGAGCAGAACTAAGGAGAGAGAGACA
CGAAAAACCCTTCATAAATCAATGAATCCAAGAGCTGTTTTTTTGAAAAGATTAACAAA
ATAGATAGATCACTAGCCAGACTAATGAAGAAGAAAAGAGAGAAGAATTGTATAGACACA
ATAAAAAATGATAAAGGGGAGATCATCACTGATCCCACAGAAATACAAACTACCATCAGA
GAATACTATAGACACCTCTATGCAAATAAACTAGAAAACCTAGAAGAAATGGATAAATTC
CTGGACACATACACCTTCCCAAGACTAAACCAGGAAGAAGTCAAATCCCTGAACAGACCA
ATAACAAGTCCTGAAATTGAGGCAGTAATTAATAGCGTTCCAATGAAAAAAAGCCCAGGA
CCAGATGGATTCACAGCCAAATTCTACAAGAGGTACAAATCAGAGCTGGTACCATTCCTT
CTGAAACTATTCCAAACAACAGAAAAAGAAAGACTCCTCCCTAACTCATTTTATGAGGCT
GGCATCATCCTGATACCAAAACCTGGCAGAGACATACACACAAAAAAGAAAATTTCAGGC
TAATATATCCCTGATTAACACCGACGCAAAAATCCTCAATAAAATACTGGCAAACCAAAT
CCAGCAGCACATCAAAAAGCTTATCCACCACGATCAAGTTGGCTTCATACCTGGCATGCA
AGGCTTGTTCAACATACGAAAATCAATAAATGTAATTCATCACAAAAACAGAACCAATGA
CAAAAACCACATGATTATCTCAATAGATGCAGAAAAGGCCTTCAACAAAATTTAACAGCC
CTTCATGCTAAAAACTCTCAATAAGCTAGGTATCGATGCAATGTATTTTAAAACAATAAG
AGCTATTTATGACAAACCCATACCCAATATCATACTGAATGGGCAAAAGCTGGAAGCATT
CCCTTTAAAAACTGGCACAAGACAAGGATGCCCTCTCTCACCACTCCTATTCAACATAGT
GTTGGAAGTTCTGGCCAGGGCAATCAGGCAAGAGAAAGAAATAGAAGGTATTCAAATAGG
AAGAGAAGAAGTCAAATTGTCTCTGTTTGTGGATGACATCATTGTATATTTAGAAAACCC
CATTGTCTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATA
CAAAATCAATGTGCAAAAATCACAAGCATTTCTATACACTAATAATAGACAAACAGAGAG

FIG. 1D

CCAAATCATGAGTGAACTCCCATTCAAAATACCTAGGAATACAACTTACAAGGGATGTGA
AGGACCTCTTCAAGGAGAACTACAAACCACTGCTAAGGAAATAAAAGAGGATACAAACAA
ATGCAAAAACATTCCATCCTCATGGATAGGAAGAATCAATATCATGACAATGGCCATACT
GCCCAAAATAATTTATAGACTCAATGCTATGTTCATCAAGCTACCACCGAATTTCTTCAC
AGAATTAGTAAAAAACTGGCCAGGCTCAGTGGCTCACGCTTGTAATCCAAGCACTTTGGG
AGGCCAAGGCAGGAGGATCAAGAGGTCAGGAGATTGAGACCATGGTGAAACCCCGTCTCT
ACTAAAAATACAAAAAATTAGCCGGGCGTGGTGGCAGGCGCCTGTAGTCCCAGCTACTTG
GAGAGGCTGAGGCAGGAGAATGGCGTGAACCCAGGAGACGGAGCTTGCAATGAGCCAAGA
TCCTGTCACTGCACTCCAGCCTGAGTGACAGAGCAAGACTCCGTCTCAAAAAACAAACAA
ACAAACAACAAAAAAAAAAAAAACTACCTTAAATTTCTTATGGAACTAAAAAAGAGCCCAT
ATAGCCAAAACAATCCTAAGCAAAAAGAACATAGCTGGAGGCATCATGCTACCTAACTTC
AAATTATGCTACAAGGCTACAGTAACCAAAACAGCATGGTATTGGTATGAAAACAGATAT
ATAGACCAATGGAACAGAACAGAGGCCTCAGAAATAACCCCAGACATCTACAACTCTCTG
ATTTTTGACAAACCTGACAAAAACAAGCAATGGGGAAAGGATTTCCTATTTAATAAATGT
TGTTGCGAAAACTGGCTAGCCATATGCAGAAAACTGAAACGGGACTCCTCCCTTACACCT
TATACAAAAATTAACTCAAGATGGATTAAAGACTTAAACGTAAGACCTAAAACCATAAG
AACCCTAGAAGAAAACCTAGGAAATACCATTCAGGCCATAGGCATGGGCAAACACTTCAT
GTCTAAAACATCAAAAGCAATGGCAAGAAAATCCCAAATTGACAAATGGGATCTAATTAA
ACTAAAGAGCTTCTGCACAGCAAAAGAAACTATCATCAGAGTGAACAGGCAACCTATAAA
ATGGGAGAAAATTTTTGCAATCTGTCCATCTGATAAAGGGCTAATATCCAGAATCTACAA
TGAACTCCAACAAATTTACAAGAAAAAAACAACCCCATCAAAAAGTGGGTGAAGGATGTG
AACAGACACCTCTCAAAAGAAGACATTTATGTGGCCAAGAAACATACAAAAAAAAGCTTA
TCATCACTGGTCATTGGAGAAATGCAAATAAAAACCACAGTGAGATACCATCTCACTCCA
GTTAGAATGGCGATCATTAAAAAGTCAGGAAACAACAGATGCTGGAGAGGATGTGGAGAA
ATAGGAACGCTTTTACACTGTTGGTGGGAGTGTAAATTAGTTCAACCATTGTGGAAGACA
GTGTGGTGATTCCTCAAGGATCTAGAACCAGAAATACCATTTGACCTAGCAATCCCATTA
CTGGGCATATACCCAAAGGATTATAAATCATTCTATGATAAACACACATGCACATGTATG
TTTATTGTGGCACTATTAACAATAGCAAAGACTTGGAACCAACCCAGATGTCCATCAATG
ATAGACTAGATTAAGAAAATGTGGCACATATACACCATGAAATACTATGCAGCCATAAAA
AAGGATGAGTTCATGTCCTTTGCAGTGACATGAATGAAGCTGGAAACCATCATTCTCAGC
AAACTATCACAAGATCAGAAAACCAAACACCACATATTCTCACTCATAAGTGGGAGTTGA
ACAATGAGAACACATGGACACAGGGAGGGGAACATCACACACCAGGGCCTGTCAGGCAGT
GGGGGGCTAGGGGAGGGATAACATTAGGAGAAATACATAATGTAGGTGACAGGTTGATGG
GTGCAGCAAACCACCGTGGCACATGTATACCTATGTAACAAACCTGCACGTTCTGCACAT
GTATCCCAGAACTTAAAGTATTAAAAAAAAAAGACCATTTATGAAAACATGACCTTACCA
AAGAACTATATAAGTCACTGGAGACCAATCCTGGAGTGACAGAAATATGTGACCTCTCAG
ATGGAGAATTCAAAATAGCTGTTGTGAGGAAATTCAACAAAATTCAAGATGACATGGCAA
AGGAATTCAGACTTCTATCAGATAAATTCAAAAAAGAAGATGAAATAATTTTTTTAAAAA
TTCATGCAGAAATTTTGGAGCTGAAAAATTCAATTGATATACAAAAGAATGCATCTTACC
AGCAGAATTGATCCTGCAGAAGAAAGAATTAGTAAATTTGAAAACACTCTATTTGAAAAT
ATACAGTCAGAGGAGACAAAAGAAGAAAATTAAAAACAATGAAGCATACCTACAGGATCT
AGAAAATAGCCTCAAAAGCATAAATCTAAGAGTTACTGGCCTTAAAAAGGAAGGAGAGAG
AGAGAGAGAGTGGGATAGGGGTAGAAAGTTTATTCAAAGGGATAACAATAGAGTATCAGT
ATTCAAATACAAGGTTATGGAACACCATTCAGATTTAACCCAAAGAAGACTACCTCAAGA
CATTTAATAACTGAACTCTCATTCAATGGGAAAAGTAAAGTCCTTTCAATAAAGGTGTTG
GGATAATTGGGTATGCAAAAAATGAATTTGGATACCTTTCTTGTGTCATATACATAAAAC
CCCAAAATAGATTAAAGACCTAAGTATAAGAGCTAAAACTATGAAACTCTTAGAAAGAAA
CACAGTAAATTTTTGTGACCTTTGATTAGGCAATGATTTCTTAAATATGATAAAATATGG
TAAAAGCAACAAAAGAAAACATGAATAAATTGGATCTTATCAAAATTTAAAACTTTTTTG
CATCGTAGAATACTATCAAGAGTATGAAAAGAAAACCTACAAAATAGGAGAACATGTTTG
GAAATCATGTATTTGTTAAGGGATTAGTATACAGAATATATATATATATATATATATATA
TATATATATATATATATATATACTCTTACACCTCAACTATAAAGAGACGAATAACCCAAT
CTAAAAAAATAGGCAATAAATAGCTATTAGTTCTCCAAAGTACATACACAAATGACCAAC
AAGTTCATCAAAAGATGCTCATCATCTTTACTCAGGAGGCAAATACAGATTAATATTACA
ATGATATTAGACATGGATTTGTCATATACAGACTTTATTAAGTTAGATTCCCTCTATGCC
TAATTTGTTGAGAGTTTTTATCATGAAGAGATGTTGCATTTTGTCAAATGCCTTTTCTGT
GTCTTTTGAGATGATCATATGGTTTTCGTCCTTTATTTTGCTGATATGATGTACCACATT

FIG. 1E

TATTGATTTGCATTTATTGAATCATCCTTCCACCCCTGGGATAAATCCCACTTGATCATG
GTGTATTATCTTTTTGATGTTTTTGGATTCACTTTGCTGATATTTTGTTGAGGATTTCT
GCATCTATAATCATTAAGGATATTGGCCTGTAGTTTTCTGTTTTTATGTTGTATTCTAGT
CTGATTTTGGTATCAGGGTAATGCTGTTCTTGTTGAGCGTGTCAGGAAGTCCAAAAGACT
TCTTCTTTAGTGTTTTGGAATAGTTTGAGAATTGTTAGTTTTTTTTTTTTATAAGTTTGG
TAGAATTCAGCAGTAAAGCCATCCAGTTCTGGGCTTTTCTTTGTTAAGAGACTTAAAACA
CACACAACGCACACACAAAATGAAATATCACTTTCCACCCATTATAATTTACAAAGTGGA
AAATAACTCGTGTTGATAAGAATGTGGAAACCTTGAAACCTTCATGCATTGCCAGTGGTA
ATGTGAAAGAATCTTGCCATTGTGGAAAACAATTTGTCAGTTCCTCAAACAGTTCAACAT
AGAGTTACTGTATGAAATAATTCAACTCCCAGGCATGCACCCAAGAGCATTGAAAACATA
AGTACACACAAAAACTTGTACAAGAACAGTCAGATCAGTATTATATATAATTGGCAAAAA
ATGGAAACAATCCAAATATTCATCAACTGCTGAATAGATAAAATGTGGCATATCCATATA
ATTAAATACTATTCAGCCACAAAAATAATAAAGTACGGATAGACACTAAAACATGGAAGA
ACCTTGAAAATATTAAGCTAAGTGAAAGACATAAGACACAAAACCCAACATTTAAAGGAA
ATTTCCAGAATTGTCAGATCCACTGAAGAAGAAACTTGAGTGTTTGCCAGCATGTGGGAG
GAGAGGAAAATCAGTAGTTATGAGGTTTCTGGAATTAGTAGTGCTGATGGTGACACAACA
TTGTGAATATACTATAAACCACTAAATGATACCTCTCAAAATGGTTAAAACATTACTGTT
GTGTTATGTGAATTTTACCTCAATTAGAAAAGAAAAAAATCTTATCAATAACAAAGAGAA
ATTTCCACACAAGGTGGGATCGCTTCCACAGTGCTACTCAATGCAGTTTAGCGATTGCAT
TTGTATTGGAGTAAAAGCATGTCACATTGCTTTTAACATTGGAGTCCAATACATAAACCT
CTTTCACCATAACTATATGGAGTTCATTGTATGTATATTTATTAAAATGGAATTAAGATG
AATTTCACAACACAATGGATCATTTTTTTTTTCATGTGGAAAATCAGAACACATGCCTTA
ATGGTTACATGCCCCACCTGCTGCTCACCTAAAAGTAAATTTCCTCTAACTCAGACAAAT
ATGTTATTTTCAAGGAAAAGAAGCCCAGAGAACTGAGATCCAGAAGAAATAACATGTATT
GAAAGCACACAGAAGTATTTCAATGAACTCAAACCCAAGATTGTAGAAAACTCTCATGTG
CCCCTGGGACTGATGTTTGAAAATACACATATTTTGCTCCTACTCTTTCCTTCCCCAGAT
CCCACCCTTCAGAGCACCCGACGATAATGGATAGTTTCTAGCAGGGTGTCTGGAATGGGC
AAGTACCCCCAAAGTTATAGTTTGTACTGCAAGACTTGAACCCACTCTTTTTCTGCCCTC
TATTATTATTTTTGCATTTTAACCATTTATTATTTTGAAAAGAAAAGAGAATTTTTAGAA
TATGGAAAGAGGAAGTGAATTAATAAAATAGCACACCCTACATAGAGACTGCTAATCCAT
CTCCAGTCTAAAGATTTAGTAATAGGCAAGAATATACATATCCAGGAATTTCCTTGGTGT
TACATAAACAAAGGCGGCACATATGTATATTTTTCACAAAATATTCACTGTTTGAAGAAG
GAATTACTCCCTTCAATTGAGTTCAGGCCTGATCAACAAGTAGTGATTGGCCAACAGCTA
AATGCAAAGTGCATGCTAAGTCTGGGGATACAAAGATGAATGAGAAAACATTTATGCCCT
TAGGAGAAAAACAAATATCTTTATCTCAGAGAATAGAGAAGGAGATTGATTCTCTTTGGG
GGAGATGTCATCCTGAAGAGTATAACAAGTTCCCCTATAATTCTACTTTTCAGTACTGTT
TAAAATACAACTGGATTTTTTTAAATATGTAAAATTTATATAATTTTACAAATGTCTTTG
TTAAGAATTAAAACTATCATTAGTAAAGGACACAGCTGGAAAATTGAAAACATTTTGGTT
CTCTACTGTGGAAACAGAATAGAGTAACAGCAAAAAGCGTATTTCTGGAATTGGACCCTG
ACAACTCTGCTTAAACACTCCACCACTTTCTAGCTATATGACCTTGGGTAAGTTACTTAA
CTTCTTTGTGTGTCAGTTTCTTCATTTGTAAAATTGGAATAATAGATGCTTTTTTTGAGA
CAGTGTCTCATTCTGTTGCCCAGGCTGGAGTGCAGTGGCGTGACCACAGCTCACTGCAGC
CTCAACCTCCTGAGTTCAAGTGATTCTCCAACTTGAGCCTCCCAGATAGCTAGGACCACA
GACACATGCCACCATGCCTGGGTAATTTTTTTTTTAAGTTTTTCATAGAAATAGTGTCTC
ACTAAGTTGCCCAACCTGGAAAATTGGAATAATAATTCATAAAATCTTCCTCCTAGATTT
GTGAAGATCAATTGAGTTAATGTATGTAACGTACTTGGCACAGAGCTTGGCCCATGTAAT
CTCTCAATGAGTGCTAACATTACTTGTCTCACAAAAAGTTACTTACTTCCGTCTGGCACC
AACTCCCTCTCTCACTTCCCACAATCTGGTTACCATTCATTCTTCAGTTCTCAGCTTAAA
CAATGTCTTTTCCATATGGTTTCATTGACGCCACTTTGGGAAAATAGATGTCTCTTCTGC
TTGCATTTTCAGACCTTTTTAGGTGTATACCTTAGGGCATTTGCTTTACTGACCAAAATT
ATTTGCCGGCTACTCTGTGCTTTTCATGACACACTGAATAAGACAGGAAGAGTGTTTATC
TATGCTCAACATAAGATAGGCATATAATGGAAGCTTCGTATATATTTGTTGAATAAAAAA
CATAAGGGGAAAATATCAGATCTAATAATGCAGGACAGGAGGCAAGATGGAACGGAGAGA
ACCTTGTCTGAGAAGAGACATAATTAAAACAGGGCATGGGAGGTAATAGAAAGATTGGAG
GAAAAAGAGACAGAGAGACAGAAATGTTTGTGGTAATTTGTGACAAGTAGCTTTGATTGT
TCATGGCCTAATCTTTTAGGGCATGAGGTTATTTCATTCTCTGTAGCCCACCGAGAGTGC
GTACAGTGACACATGTTATGTAAGTCCCCTTTTCCCTTTTTATAAATGTCTAGACCCCCT

FIG. 1F

GTGATTTGAGACTTTTCTAGAAGAATTTAGCTGAAGACCATATTGTTTTTTAAATGTAGT
ATTTGGAGCCTAGAGGTGCCAGATAACTTCCTGCAAAGCTAATGCATTTATTTTGGGAAT
ATATAAGCTCAGTATCATCATTACCAACAGTGCTCAGACTTGATTTTATTTTCATTCCAA
CAGCAAAGGAAAGAAAGCAACTTCTTTCATGCTTCCATGCCACTCTGCATCTCTCTACCT
TCACAGAGTTTCTCAATAATGGCAACATTTCCAGTTCACCAATGGACTGAGAGATCATTG
AGGCTAGACTAGTCTTATTAATCCTTATACCCCAGCTCCTAGCCGAACTCCTGGACACAC
AATAGATACTCAGATACATTTACTGAAATGCATATAGAAAGTTACACCTGCAAAAAAGAT
GATCTCTCACCAGGAATAAGAAAATATAATCTGGGACAGCCCATATATGAGATCTCTAAA
CAACCTACCTATAACCACCAAGAAAAAAAAATACCTGAGTTTGAGATTTATTTTTCCGTC
TCATTTTTAATATATTCCAGTTAGTGAAAGAGCTAAAATAAATGACAAGAAAAATTTAAT
CTAGGTATTTAAACAGAATTATTCTGAATGTTGTGAGCTACATTTCTTTTTTACCTTTTA
TTTATACATAGTATTTGTATATACTTATACAATATATTTATTTTGTATATATAAATATAT
TGTATTTATTTATACATGTAAATGTATAATATATTTATTTATACATAGTATTTATATATA
CATAGTATTTGTATATATTTATAGGGTACATGTAATATTTTGTTACACGCATAGAATGTG
TAATGGTCAAGCCAGAATATTTAGAGTATCCATTACCTTAAGTATTTATTATTTCTCTGT
GCTAGGAGCATGTTAAGTCCTCTCTTTTAGCTATTTTGAAATGTACATTGATGTTAACTA
TCATTAACACAGAGTAATTGATATGTATAGCAAATAATATTTGCAGTAGGATATCACATG
TTTACTTATTTATTTATTTATTTATTTTTATTATACTTTAAGTTCTAGGGTACATGTGCA
CAACGTGCAGGTTTGTTACATATGTATGCATGCGCCATGTTGGTGTGCTGCACCCATTAA
CTCCTCATTTACATTAGGTATATCTCCTAATGCTATCCCTCCCCCCTCCCCCACCCCACG
ACAGGTTCCAGTGTGTGATGTTCCCCTTCCTGTGTCCAGGTGTTCTCATTGTTTAATTCC
CACCTATGAGTGAGAACATACGGTGTTTGGTTTTTTGTCCTTGCGATAGTTTGCTGAGAA
TCATGGTTTCCAGCTTCATCCATGTCTCTGCAAAGGACATGAACTCATCCTTTTTTTGGC
TGCATAGTATTCCATGGTGTATATGTGCCATATTTTCTTAATCCAGTCTATCATTGTTGG
ACATTTGGGTTGGTTCCAAGTCTTTGCTATTGTGAATAGTGCCGCAATAAACATATGTGT
GCATGTGTCTTTATAGCAGCATGATTTATAATCCTTTGTGTATATACCCAGTAATGGGAT
GGCTGGGTCAAATGGTATTTCTAGTTCTAGATCCTTGAGGAATTGCCACACTCTCTTCCA
CAATGATTGAACTAGTTTACACTCCCACCAACAGTGCAAAAGTGTTCCTATTTCTCCACA
TCCTCTCCAGCACCTGTTGTTTCCTGACTTTTTAATGATCGCCATTCTAACTGGAGTGAG
GCACTGGTCTGAAAATATCAATTCATTTAATTCTTTTAACAACCTTAAGGGGATATCATG
GTACAAATTTAGAGCTTTCTTTTGTGTTTGTAAAATGGATTGATTCCTTTTCCCTACATC
CAGCAGAAATATTTGAATTGAAGAGAAGAGTAATACCTAAGAACTAGAAATTCCTTTCTT
ATGTTTCAAAAGATATCAAAAGATCTAAGGAAGATATTCACATCAAAAATGAGTATTATA
ATATTTATTATCTATGGTGCACTTGCAAAAAAGAAAACAAGTAATAATCTGAAGATTTAA
GTGAATATTTTATGACATTGGAGTACCACATATTTAGAAGAAAGCACCAGAGAAATCATA
GATAGAAGGAAATGGAATATTTGTAGGATCAAGATAAATACAGCTTGTCATAAAATAAAG
CAGGTATCAGGATAAAATCTTGAAAATATTTTCATTTCTCGTTATTTATAACTTCAATTT
ACTGTGATGATTAATTGTAGGTGGAAGATTTACGAAGAGAAGACTGAAGTATAGACAAGT
TGAAGTGCCACAAAATGAAAGCTAATGACACTGACTACTTAGGAAATAGCAGACTGGGTC
CATATTTATAGATTGTCAATGACAAGGAATTTGCAGATGTTAATGAATATAGATCCGAAC
TTAAGTTGCAACAACCTTTCCCACTTTGAGATGAATAGTGCATGGAAGAGTAAAATGCAG
ATGTTAATAAATCAGAGGAAGACATCGTGCCAGAGTATAAAGTTGACAGATTTATGCCGA
TGAACTTGAACAAAGCCACAGAAGGCCTACTTGTCAAATTTACTGGTGACAACAGGTCTG
GAGAAATGGCTAATGTTTTGGATAATAGCATTAGAATTTAAGGTCTGTTTAAACTTCAAA
TTAACAGAATGAAATTAATATATGCACATATCAATTGGGTCTTTTGCTTATATATCATCT
CTTAATAGAGCCTTTTTGAACAATCATTTCTAATGTGACCTTTGGGATTTTCTACTCATC
ATCACCTCATCCTGTTTGGTTTGCATTATAGCATCTATCCCTTCCTAACGTTTTCCCTAT
GTATTTGTTAGTTTGTTTTTTTTTAATCTAACTTTACTAGAAAGTAAAATGCATGGAAAC
AGCAACCTGTTTAACTTTGTATCACTAAGAGTGGAAAAATAACCCTCAGGAAATATTTGG
TAAAATAATAAAATGCCCATTGATGCCCTTCTCTTAAAAAGAAATTTAATTAGTGCAGAT
TGGGGAAATACAACAATATTTCTCATAAAATGTGATATCTATACAATAACAGAAGTACTA
TGTCCCAAAAAGTATTCTATAAATAGAAGAAAGAACAGATGGTTTTGCTGCTGATTAATC
CATTTATCTTTCGTAAATCATCTAATTTCCCCAGGAACAGCTTCCTCATCTATTAAAGGG
GGTTAGTAATAGCTAAGCCCTCAGGGGTTTAAAAATGCATATGAAATAATTTTATAAACC
ATAAAGCACAAAACAAATATGAAAAATTATGATTGGAGGAGGGGGTGGGGTAGTTAACTA
AATCTCAGTGTAAACCACCAATGTCTTGTGTGTGTTGAAAAAATAATTACATATAAAAAC
TGGTTGCATCCAAAGAATAATGTACTTTTTGCACTGGCAAGACTCAAACCATATTATTGT

FIG. 1G

TACTTCCTCCCAGTTACATATTTTGCAAGATATTGACAATTGTCTAAAGGAAGACCAAAC
AGATGTAGGTGGGAGCTACTGTCATTTGAACAACATTGAAAAGAAAAATACTAAAAAAGA
AACATGAGGGCATATAAAGGAGCGCTGGGGCTGTGATGTTTATTTTGAATCTGTGAAGCA
TTGTCATGTGGAAGATTTATTCTGTGTAGCACCAAGATGCAAACTAGGAATTAGAGGTAA
AAGTCTCAAAAAGACAAATCGTGGCTTGAGACCTTGGTTTAATGTAAGAAACAGTTTTCT
CACCCTTAGAGCACTCCCATAAGGATGGAAGTAGTGAATTGTGGTGGTCACATTCAAGCT
AGATGGGGACATGTCAGCAATGTTATCAGGAGGCTTCTACTCTGAAGCTGAAGTTCAGAC
AAGATTTCCAGGCTCTTCCCAAGTGCAAGATTGTAATTACTTAAATGCAATATTTTTACC
ATGTTTATTAAGAATAAAAGGATCATGAATTCACATTCTGACAAATGCTAGAATACTTAT
TATTAGAGACAAAACCAGTGCATGAGAGAATGGCAGGTGACATCAGCCCTGAATCAATGG
GAAGAAAGACCCAATGGGATGTGGTATTTACCAGAGAGAGCACTTCTGCTTAGATTGCTA
CATCCTACAGTGAATGTTTAATATCATTGAGTATATTGGTGGTCTGTCATGCTTGACAAC
ATTAACTATGATCATATTTATGACACTTGGCGTCCTTCAAGAATTTGTAGCTCTATTTCA
CATGACACTTAACTATCGCAAATACAAATTCCAGCTAAATAGACCCTTCAGTTTAAAAAC
AGTCTCATTCTCAAATTTTAAGGAGAAAGTGAAGACGGAGATGTCTTAAAGACTCGGCAA
GTACTAAGTTGGCAAATGTCAAATGTTAAAATAAGTTTATATTAAATGTTAAAGTGTTTG
CCTGGAATGACTTTTCCATTGTCCTGCTTGAGAAACACAGAGGCACCTCCTTATTGCTTT
TATATTTGCTTTACAAAGACAAATGTATCAACATGCTCTGTATTAATTGTATGTTGACAT
TTTTGTCATATCCACAGACTGATGCATGTCTGTGCATGGTTTATAATAAGTGCACGTAAA
AATAGAGAAAATAAGTAGAAAAAGAGAGAGATTTAACTCTCACCCCCCACCCCCCAAAAA
AACAGATTAAATTAGTTTTCATTACTTTTTTTTTTTTCTTCAGCTTCAGCTCTCCCTCAG
CGAGGGAGGAGGCTGTGGGCTGCGGACTGAGTGCTGGAATGAGGAGTAATTGAGCTTCAG
CTGAGCCGGACGTAGCTTTCTCCTCCTGGTGTCATTGCTGCAGCCTCCAGTGCCGGGTCC
CTAGTTCCTCAGCTGCCTATCTTCCCGGTGCAACATCGCCTGTAAAGACAGCAAAGCCAC
CGCAGAAGTTGCCCGGCAGAAGACTCCGGAGGCATTGGCTCAGTAACTTTTCACGTCATT
TTCTGCTCGGGAGCCCCTTCTAGCCTCTCCGCGCAGCCTTTCCCACCGCAAATCACCAGT
GCTCATGGGGCAGGCGGAGAGGAGCTTGCAGCATTGAGCGGAACCGGACTTGAGCCCGTG
ATGTCCGGCACCAAATTGGAGGACTCCCCCCCTTGTCGCAACTGGTCATCTGCTTCGGAG
CTGAATGAAACTCAAGAGCCCTTTTTAAACCCCACCGACTATGACGACGAGGAATTCCTG
CGGTACCTGTGGAGGGAATACCTGCACCCGAAAGAATATGAGTGGGTCCTGATCGCCGGG
TACATCATCGTGTTCGTCGTGGCTCTCATTGGGAACGTCCTGGGTGAGTCTCCTCCCGGG
CAGCCCTCCTAGGGGCTATCACCCCCTCTCCGCCCCGGGCTGAGAAGGCTCTAAAGAGAC
CCCTCCCTCCCCCGGGAAGCAAACAAAGAGGTCGCTGCTCTCGGATGGGGTTTTCTAATA
AAATAATAATAATAATAGAAAGTTTTCTGATTTTCCGAACCGGGACCGAGCCCTGGAAAG
GTTATTCCCTGTTTTGCAGGAATAACGGGGAAACCGCGTTTCTTTTTCGAGCACCTAGAT
TACAAGCGCAGGGAGAGGGGCCGCGGCAGGGATCTCCAGGTGGATTTTGTTGAGTGTGTG
TGTGTGTGGGTGGGTAGGTGGGGGAGTCAGTCATCCCTTTGTGTAACGTGGCTGGGTGTT
TCAGGGGGGTTGGGACGAGACAGAGCTTGCAGAATACAAAGCTACATCCCTAAGGAGCAA
GCTCTCTGTGGCTGTGGAAGTCACAAAGCATTTGTGAGCTAGGTGGCATTGCCCTTTGGC
GAGGAGGTTTAGTCTCCAGTCAAGAGGTGGTAATGAACCAGCAGGGAGTGGAGACGGAGG
CAAAGCAGGGAAGTGCACTCACTCATAGAAGCTGAATTAAACAGGATCCATGCCTGGAGC
AAGAAGGAGGGGCATCGGAGAAAAGTACCACAGAGATCTCAATCATCCATCCATCCATTC
ATTCTTACATCCATTCAGCCAAATATTTTTTTTTTTCAGTCTGCTTGTTGCCAGGCTCAG
GAATTATTCATGTCAACTGTTTGTTGTTGTTTTGTTTTGTTTTGTTTTCTCCAAAGATGA
GACTAAGCTTAATGCTAGGCTATTTGTCCCGGTCTAGGTCTGTATGCAAACACGGGTTTC
CTCGACCCCTCATCCCCCTCCCCCTAAACAATTTCTGAGGGTTGGGGAGGGGGTGAGATG
GCAACATGGTGAGTGCGATGATGGAATGTATTAGGGCAGTTGGGGAATATACCTCCAGAA
AAGGGGCTTTGGAAGGGAGGGATAACTTGAAATAAATTGTGAATGGAAGGAGAGTGTACC
TTGATGAATGAAGAGTAGAAGGCTGGGAGACTTTTCACATGCAGAGGGCAGTGTGGAGGA
AGTCTCTGCTGAAAATGACAGGAGATGGAGGAGGCTAGGAGTTGCTCTTGATTTTCATTT
ATAAAAGAAGAAGAAGGTGAGTGAGGTGAGATAGGCTGGGAGGCTTTGCAGTCAAAAGCA
AAGAACTTGTAGCTGCAATGGGGACTGACAAGGAAATTATCAGGCTTTCAGACTAACCTG
ATTTTTGCCTTCTCTCCCAAGTGTGTTGGTCTGGGTAGAAATCATCCCGAGTAGTCTCTC
ACCAACTCAGCAGGCAGAATAGATGATAGTATGTGAATGACAGGAGTTCTCCAGAGTGTT
GGTAGAATGTTATTTGAGGAGACAAGAAACCTCTGAGAACTTTAGTACATTTTTAAATAT
TATTTTTAGACTGTTTTCCTTTGGTTGATTTAAAAGTAAAAATAAAGGAAATCTTTTTGG
GATACTAACAAAATGAAACAAAAGTGGAAATACACAAGATTAGGATTCTTGTTATAAGCA

FIG. 1H

TAATTCTGTTGATAATAATCCTAATCTTGCTTTCCTTCTTCTTGTTACCCATCCTTAGGA
TTACATCTCTTAAGACACATGGCTACCAGCATAGCAACATTTTACTGCATTATGCCAACA
CTTATTGATAAGTGAATAATCAAAATTGAACATATATTGAGTACCTACTGTGTGCCAGAG
CCCTTCATGTACATTCTCTCCCTTAAATATCAAAATAACCCACATTAGCCAGAAGAAGAA
ACAAGACTTAGAGAAATAAAATGACGTATTAAGGGACATAATTTAAATTCAGTTCCATTT
TTTCTGACCTCAGATCCAGAATTCTCCATTGTTATTCCACTCTAGAGCTAAAAAGCATAT
AGAGAATAGATTCTCTGCTCCTGATTGTCTGCAAGTTTATTAGATGTGTTCCTGTTCTCC
TCTGCATCAACGCCCACTGCCAATAAAGTACAATGAGGGATTAATGGCACTGTCATTCTC
TTCACCAAAAACCTTTCCAGAGAAGCAGTAATTTTTTATGAATAGCTATCAATAGTAAC
TATTTGCCTTCCTTATTTTAATTTTCGGCTGAATCTTTGTGGTAAAATGTGCTCTTCTTT
GTTGTTATTGCATTTTTACCTTGCATAGACCTTGTAGTGAATAGTCTCCATATCCTAATT
GCATAGTTTAGGGATACATGTTTGCTAGCCTGGGGAGTTTTAGTTTCAAGAAGGAAACAC
CTCTACAGTAAGGCTACTTGTTTCATAATGTCAAGGAAGATAGCACTGTCCACAGCCCCA
AGTGCTGAAATGGCCAATTCCATTCAGCCTAAAAAGAAAGATTTACTCAAAGCACTCTGC
CTTAAAGAAACTGACAGCTATTTTCCTCAGGACTGAATAACACTGAAATCCTCTCTGGTT
GAACTGAAATGCATTCTTTTCTGACATACTGCCTGAAAGTTGATGAGGTTTAGGTTTGAC
ATTTAAACAAACGAGTAGTGTCGTTACTCACAGACAACTTCCTGCTCTTTGATGTCACTG
TCAAATTTGCAAAATGAATTAGATTGAGAATTGCTTCTTTGCCCCTCTGGTATAAGTAAT
TTTGCACATAGAGTGGTAGGACAGGATGTCACATGATTTATGCAAAATAAAGATGCAATA
TTAAGTATGAAGGTAAAATACCACAGTGTAGGCAGCAGATGTAATCACTGAGCCTTCAGG
TCCAGTCACCATTTGTACTTTCATATAACTGCTTGGAAAATCTCAACCTTTTTGGGCTTA
CAAATATAATGCCATCAGTTAGAAGTCATCTTCTCCACAATGTCCTTTCATGAAGTGATG
TAATAGGATATGCTGTGGGTAGCATAACAAAGTCTTGATTGTCCTCATCTCTTTTTCTTC
TCCCCATAGTCCCTCTTTATCACTATGCCACCTCTCCACTCTCATATACTCCTCCCAAAG
ATGGAAAGCAGTTTCCTGGGGGAGTAAAGTTTTAAATAGAATGTTATGAGTATTTACATT
CAATGAAAAGCTGTAAGCATGTTTAATGTGAAATTTTAAGTTCTAAGGAAGGAGCATAGG
GTAAGGTTCTTTTTGGAAGGAGTATCTTTTCAGTATCTTCAGAATAATGCCACCTATAAC
CTATTCCTAACTATGTCTTCTACTACAGCTAAGTAGATGTATCAACTTATTCAATTGGTA
TATTGTGAGCATTATCATTTTTTTAAATTAGTGTGTATATCAGGGGAGCCTCTGGGGAAA
TGTAAAGAAATGTGACTGATGTTAATTTTTACTCCTGATTCCTTGAATGACAATTGTAGG
GAGAAATGTGTTCTAGTCAGTTTAAACATTAAGTACCTAGGGAAAATGATCAATTTTCTG
CTTCTCATATCTGCATTCAAAGATATCATATGTTTCATCTGGTATGCTTCTGTCATATCT
GTTGTTGTCTCCATATGGAAAATAGGAAAACATCAGTCTAGCTATGCTTCTTGCTTCTTG
TGTGCCATTAGCAAGTTATTGAACTATCCAAGTCAATTTTTTTATAATTACAAATTAAAG
ATCGATAATGACTGCATTATAGAAATAGTATCAGGATATAATGTACGTATACCCTCTATA
AAGACATATAAAGGGACACAGGCATATACATATTTTTCTTGACACATAGACACTAATTAA
TGTCAATTTTTATCCCTTAATTTTCATGACTGAACTTTTTGTGATGTGGTGTATAGCCAG
CTTCTGCCTTCATGGGCCAGTCTGTATCTCTGTAGCTCTTTATGGCCTCTGCCCCAGCCT
TTTCCTTAATTGCATATTTTCCTAAAAGGTGTGAATAAAATGGTGTTGGCACACATTACT
CTCCTTTTCCACACTAGCTCCACCCACCCATCTCCTTCATACTGATTGCTTAACATTGCC
TTCTTGCCTTTAAATGAAAGCCATTCCTAACTATTGGAATAGTTTGCTTTCTCTCTCAAC
TTAAATTTGCCTGTGCTGGGTCCCATTCATTTAGAGTTTTTGAGTTGTTAATAGGTTGTT
GATAGGCAGGTCTATCACTACTAGTGTTTTAAATAACACACACATTGGTAATATGTTGAT
TTAACTCATACATTGTTAAAATACATTGTGAAGTATTCATAGTTAAAATAAATTATCCAT
TAAGTAATTTACCTAATAACAGTTTACCCAAGTTAGGTGTGTGGAATGGGGAAATATTTG
TAATAAGTTTGCTTCCTACAGAGTTAGTCTTGTGTCAGATATGTAAGTGGTAGAATTGCA
AGTTCATGTTACTCCTAAGCCTAGAGACATTTATTTTCTGCTTCTCCGAATGCCCATTTT
AGTTTCATGGGTGTTTGTAAACCCATCCTTACCTACACAGGAAGCAAAAAGGGGTTATTT
CTAAACCCTTTTTAGATATAGAAATAATACATCACTCATCTCGGCCAAGACTCAATAGAA
TCATGAATAGTGACTGTAAAAGGTAATATTAACTATTAGGCTTTAAACCTATTGTGCATT
TTAGTTTTAAAATGCAAACATGCTAATCTGAATAAGAATTAATCTGATGCCTCTACATTT
TTGCTAAAATCATAACTGTTTAGTCTTACTTAGTAAAATAAATTATATCTTTGACTTAAA
ATCCCAATGATAACTTTTAAGATGGCTATTTCATAGATAACAGCAACATTTATCATGGAC
AGACAATAATGAGAATAACATGTGCAACTGATAATTTAAATGCAATGAGTTATTTCTGTA
TTTGAAAAAATATATTTGGGAAATGGGATAATTAAAAAATACCAGTTTTCAAGAGACCAA
ATCTAAAACTCAAACATAAACACAATGCTCCAGTTTTTAGAAAACTGTCTTGATTGTAGT
AGTGCCTACATACTAAATTGTATCATATGATTTATATTAATTTTCCTTATTTTGTATTTT

FIG. 1I

AGATTATATTTGAAAATTTTCATGTACTGCAGCTATGTTAGCATCTCAAAGTCTCCATAT
TCTCACTCCGCTCCGAAACATCCACTGCTGATGTTATTTAACTAGTGAAAGAAGATCCTT
CCATGTTTCTTCTTATAGCATTCTGACATCTTCTCCACCCTAAGGAATGCTGGCTTTATT
AAGTATGTTTCAGTCAATGACATGTGATTGGTGAAGCTGACGGTATTTGTCTTCAGTTCC
TTTTTTCCCTGCAAAGGAAATTTGTTGAATATTTATTGGGTACTATATGCCAGGTACTAT
ATGTCAGGCTCCACTTACATATACTCTATTGATGCCTTACAACAAACTTATAATGAGAAG
ATTAATAGGTTTTACAAATAAGAAAAATGAATTCAAAGAGCAATGCTAACTTACTCAAAA
GTTTAGTCAGGCAGTAAATAGCAGCACTAGGTTTCAAATATGGATTTAACAAATTCCATG
GTCCATGCTTATTCCATTACTTCATCCTGCCTCTTTCCTTAGCTTCTAACCCTGACTGGA
GATGCATAGGCAAAAAGAGGAAGGAAGAGATACTTAGATGTGCCCTCTAGACAATTTACA
GAGTTGTTTGGGCATGTTGCCATGCTGTTTTTCTGATAGACTACAGTTCTTCAGCTCTGA
GGATGAGCTCATTTGATAAGCCAATCAAGGTCGGGCTAGGGTTACTTTACAAGAGAAAAT
TTCAAGGTAAAATAGGTGCTGCCAAAAATGCTTTTACCTGTTCAGGGGGTTGACTCACTG
GAAAAAAAATGTTAGATAATTGTGGCCAAGGATTATTTTGTTATTGAAAGTGCTATTTTT
AGACACAATTTGAGCCTGAGAGCCTAAACACTTAACACTTCACATAATCTACAGATATTT
GTTTATTTTTCTTTTTGTCATGCATTGCCAAATAAATAGTATTTATTTAAACAAATCATG
TTGCTATTGATTTTATTAAATAGATGAACTTTTTTTAATTTTTTTTTTTTGAGATGGAGT
CTTGCTCTGTCACCCAGACTGGGGTGCAGTGGCACAATCTCGTCTCACTGCTGCCTCCAC
CTCCTGGCTTCAAGCTATTTTCCTGCCTCAGCCTCCCCAGTAGCTGGGATTACAGGCACA
TGCCACCATGCCCAGCTATTTTTTTTTTTTTTTGTATTTTTAGTAGAGATGGGTTTCACC
ATCTTGGCTAGGCTGGTCTTGAACTCTTGCCCTTGTTATCTACCCACCTCAGCCTCCCAA
AATGCTGGGATTGCAGGCATGAGCCACTGTGCCTGACGTGAACAGGTCAATTTCTATATC
ACCGGACAGTGTTCCTGGATCAGAATAATATATTATATGTATGAAGAATCATTACCTATT
ACATCAGACATGAAATGACCTTTAGATACTGACTTTGAAAGAGTTTGAGATGCTATTGGA
TGAAACACATGACCCATATGACCAGTCTTTTGAATTGCTGACTCTGAGTATAAAATGTTT
TCATTTCACCTTTGTTCACAATGAGAAGTGATCTCTTAACCAAGTAAATGAATTAAATCG
ATATTTAAAATAACATTAAATTTCTTGCCAGAAAAACTGTTCTTTCATAAACAAAAAACA
AATTGCTCAAAATAAATGACTATATCTTTATTTCTAAAAAATGTTTAGAGATTATTATTA
TTGGGTCTTTACAAGTAATTTGCCTTCAATACTAAACACATGAGAACAATGTTTAATATT
TATATAGTATTTTACTCTTCAGAAGATATTTGTCCATATTCTCTCTCAGTTATTCTTCAC
AACAACATTATGAGGTAGGTCTTTTTTAATGAAAAAAAACTCAAGTGCTTGAAGTGATTT
AAAATCACTGTGGAAGAAAAGCATGGGCATACAGAAAAGCCAAGTGGTTGTGTGTCAGCT
TGGGAAAAGCTTGCAAATTTCCTGTATTTCAAGAGGCCAGGATGAGGTGTGTAATTATCT
TTTACTGGTCTTCAGCTATCCTGTCTTTGATATGTGATTGTGTCAAAACTATGAGGAAAA
ACTCACATTAACAAACTTCATAAACTTGTTAAACATAAAATAATAATTTCGATGTTTTAA
TTTACAGTAAGAGTTTATTCTTACAAGTCCTTAAATACCCAAAGTTCTTTCAGTTATCAT
AGTCTTTTTCAGTAGACAGAAATCCATGTGGACTGTTATTGTTCTGAATAGCTAGGCTAT
GCCATAGTAGCAAACAAACCCTGAATTTTCATTGGCTTAGTATCACGAAAGTTTATTTCT
TGCTCATTTAACATCTGAGGTGGGTTGGAGAGTCTCCTTCATCCAATGACTCACAGTTCA
GGCAGCCTCCACATTTTGTGCACTATCCCTAAAAGGTGGACTCTGTGGTAATCAGTTTCC
AATATGGCTTCCAATGACCGCCCCCGGGCCCCGGCCCCACTTCCTGATAGTCACATCATC
GTGTAGTCCCTTTGCATATTATGCCAGAATTGGTCTGGGTGACCAACAGCTCATAGCAGC
AGTGAAACGATGTCACTTTCAAGATTACATAACAGGAGCTTACAGCTTCTGGCTCAAGTA
CCCACTTTCTCTCTAGCTCTTGGATCTCTTCTTCTGGAGGAAGTAAGCTGCCTTGTGGTG
AGCAGCTGTTGGCTGGAGTTAAAATCTCCAGCCAGCAGCCAGAGAGGAAATACGGTCTGT
TAACAACCTCATGTGTGAGCTTGGAAGCAAATCCTTCAGACCAGGTTGAGTCTTGAGGTG
ACTACAACAGCCACTACCCCAACCCACCCCCAGCTTCAGTGCAACTTAGTAACAGACACT
GAGTCAGAACTATTCAGCTAAGCTTCTTGCAGATTCCTGACCATTCAGAAGCTATGTCAT
AATAAATTTTTGTTGTTTGACTTCAGTTTCGGGATAAGTTGTTGCACAGCCTCTAAAGTT
GTGAACTAGAAGAAGTATACTGGCTCTTAACCACCTTTGCCAAAAATTAACACTTGTCAG
TCATGGTCATATTCATTTGGTCCAAATCAATCATATCGTATCAACCTAACTACAAAGGGG
ATTGGGAGATGGTGATGTCTCTGTCACAGAATCTATATAATAGTTAAAAGTATTTTTAAC
TTGCATAGACTCAGAACAAGATAATTTGGAGGAATTCAATGCTTAATGGCATACCACTAA
GATAAGCTGATAGATATATCGTTGCGATTTGGGTCTCTGACAATAGAGGCAATTGATAAT
ATTAAGAGACTATGTGCCAATTATTGTGCTTGGATTGAGGGTACAAAGGTAATAGAATCC
AAGGAACCTGCACTCTTTTTGAAAGATAGACACATAAACACATACTTTTAAAATAACGTG
GTAAGTGCTACTATGACAGATGGTTGCACAGAATGTAGTGGAAGTATTTGAGAAGGACAC

FIG. 1J

TTAGCTCTGCTGGGGGATTAGAGAGAGATACAGGAGGAGATGACACCTAAACTGAGTTTT
AATAGATGAATTCAAGTTACCCAGGTGAAGAAAATTGGGTAAGGATGTTCTAAGCAGAGG
AAACAACATAAGCAAAATCAAAGAGGCGTGAAATAGAATGAGCTATGAAGAAAGTGTTAG
GCAATTGGGTAAGTCCAATGTAAGTGCAGATGAGGAGAGTCTGGAAATGAGGCTGAAGCA
GTAAATAAGGATTGGCCATAAAAGACCTTGTGTACAATTCTTAAGATCTAGGCTTTGACA
CTGTTGTTTAGGGGGAGCTGTTAAAGGATTTTAAATTAGAGTACCATCATTGGTTTGCAT
TTTCCATGAGAGCATTTTGAGGAAAATGCAGAGAATAAATACATGAGGGGAAAGACTAGT
GAAGGTTTTCACACTGGGGTTTGCATCCTGTTTTGGCAATAAGCTTGTTTTAATGAAAAC
AAACAAACAAACTGACAATAAAGAACATAATCCAAATTCTCCAGATAATTACTTCCAGGA
GGCTTTCTACGTGCTGCATACAAAACAAAGAAAGAAAAACATAAAGTGAGAAAACGAAGG
AAAAACAAGGAAAGAAGAGAAAGAAAGAATACATATTGGAAAAACTGTTGCTGTTTTTGT
TTTGCTGAATATTTAAATTTGAGAAGCAATTTCTCTTTTTCTTTTTACTTTTTTTTTTGA
GATAAAGTCTCACTCTGTTGCCCAGGCTGGAGTGCAGTGGCGCCATTTCAGCTCACTGCA
ACCTCCGCCTTCCAGGTCCCAGTGATTCTCCTGCCTCAGCCTCCCCAGTAGCTGGGACTT
CAGACATGCACCATCACGAGCAGCTAATTTTTTGAATTCTTAGTAGAGATGGGATTTCAC
CGTGTTGCTCAGACTGATCTTTAACTCCTGAGCACAGGCAATCCGCCCACCTTGGCCTCC
CAAAGTGCTAGGATTACAGGCGAGAGCCACTGCACCCAGGCGCAGGTTTTCTTTATGATG
TTTTAATTATATCTTTCTTGGAACATATATGTATGAATCTTGCATGCCATAGGTCTATTA
ATATTTTCCAATATTCTACATGGTTTTTTACTAAAATCATTTTTATGATTAGTTACTGAC
TGAGGTTTCAATGCATCACTGTACTCCTAGCTATCTCTCATTTTAGCTTTTACATCACAT
TTTGGCCTCACACTGAAACACAAAATATTAAAAATTTGAGATCTAATAAACAATTTTCAC
ATTTTCCAACTAAATCCCCACTTCTTTCTAAATTTTCTACAACTTTCTAAACATTCTCAC
TTGAAAATTTATTTTAAATGACATGTATTTTATTCAAACAATCAATGAAGATGCTACATT
GACCCCAAGTGAGCCCTTAGGGAATTTCCGTGAATATTTCCCTACAGGTTGGCATGGTAA
CACACTTCACAATTTCTAAATCTGTGGATAGTTTAGAAGCTTTTATTTGCTGTTCCTAGT
TCACAATGGAAATACAACAATGATTAAAAATTATAATATCCTTTTGTAGATTCTTAGCTT
TTATTCCTACTCAGTGACTCTAAAATGAATTTATAAGGCCCATGGTTTATAACCATGTGA
GGCCTTGATTTTGTCACTACATTGCTAGAAATGGGGTCAGAAGGCCACCAGCTTTAATAA
TTTAATTCATCAATTCGGAATGAATTTGATGAGTCAACCACTTTGGTAGAGAACCATATT
GCTCATAAATACTGTTTTGAAGGCAATTCGTCTTTCATAAAATGTGAAGATTGTGCTGAT
CTTTCTGGGCAGGGTTATGGAGGTGTGATTAAATGCTTAAGAAACCATTTTGTTATTATA
TTAAACCGAATCAACTTTTTATTATTAAAAATAGATAAAAACTTAGCATCCTCAATTATA
ATACTTTATACAAAAGTTTCCCAATTTTATATAGACTGAAGATAAAAATACATTAACAAA
TCTTACCAGCTGGTTCAGGAAAATAACTTCATAATTATTGAGACATTTATGTGTTTGGGC
TTGATTTATACTTTGGACACAGGAAAACCTAGAGAGATCTGGTTCTTTGAAATCATCAGA
GATGGTGATGGTGACTCAGAGATTCCTGAAAATCAGTAAGATTACCCTAGTTTATAGACG
TATGTGTTATTTTTTCCCCCAGGCATAATGAACTTTATAACTTGTCATTGACAAGAAGCC
AAATCATCTTAGAGAAAAGGGGGAGAATAAAAATTTAAGAACTTAAAAACACATAAATAA
AAACATGTACATACCTCACACATGTGTACACACACAGTTTGGGGATTGGATGATATGAAT
AATATAATTAATACACCCTAATTTTTCATGCAGGATTAAGAAAGTATCTTCCAAACATTA
AAAATGCTGAAAACTGGACATAAGGCCTTGAGTTTCCCAAATTCAGGACATATTTTCAAC
TATCCCCTGAGTAAATGAACTATAACATTTACAGAAGTAAAAATGATAAATACACTAAAG
ATGAATAAGTCCTTGAATTAACAGCCAAACAAGAAGGCGCATCCTTTGGATGATTGATCA
CTGTAGCATGATTTCTTTTCCTTGAATAGACAATATTCCTTGACAATCTTTCTGTAAACA
GAATACAATGTTTCCCTAAGCAATATATGCGTGCTCTAGAGTTTTCACAATTTCTGATCC
TCCTATGACTGGCTCCTGCTCAGCTCACACTGCACTTTCATGGAAGTTCTCTTAGAATGC
CAGCTTTGAATCACTGCTCCCTCATGTGCTGTGTGTGATAGCATCCCATTTTAGTTTTGT
CATAGAATTGATTACCATTTCAAATTGAATTGTTAATTTATTGTTCATTTTTCTGTTGTC
TCCCTTAAGTAAAAGGTAAGCTGCATGAGAATAGTTTCATTTTTTTTCCTGTTTGCCAAT
GTATCCTCAGTGCCGAGAACAGGTTCAGGAATACAGAATTTTTAGTTAGCAAATGAATTA
AAGTGTAAGACTTCCAGCAGGAGGAATTTTTTACATATAAGTACATTTTTTAAATTAAGC
ATTGCAGGCTTTAAATTTCTTCTATATAAATATTTAAAATAAAGCTTCAATAATTTGAAT
TGCTTTTGTGATTATTTTGTTTTATACCTTGAGTAACTTATACATCAACTATTTTGTAGT
TATTCTAGTAATGATTATGAAAGACCATTTGAAAATCTTTCCCCAGCACTGAGATCTCCT
TGACATGACTAAGTGATTTATACTATGCAATTATATTGCTCTTCTCAAGAAAAGCAAAAT
GAAATTTACAAATTTGGTAGCTTTTTGTTCTTTTGTTTTCTCAAGTAAGATACACCAAGA
TTTCTTTAAATGATACGCTATATTTCTGCAATAACTGAGAAGAACATGTAATGTGCAAAA

FIG. 1K

CTCTTAAACTCTTTTGTTTCAAAATAATTCTTGGTTGTTTTATAAAAGTCTAAGCAAA
TACTTAATGAACTGTGTCCCAAATGAGGTGAAACAGCTGTGACAGAATGTTACTATGACT
CTGTACTTTCTATAATAAAAAGGGACAGACATATCCTCACCTGAGCCTTGGGATGTTTCA
GGCATGCCCATAGAGCCTAAGCTTTAGGAATCCTCTGTCATTCTTTTCCATTGCCAGTGA
CTTGTGCCAATTCTAGGGTTCTGGACTGTGCAAACAATGGAAAAAATAATAACACTTTCA
GGTGGCGCACAAAACCAATGTTCATAGTAGATGGATAGTTCTAGACACTTTATTTAATAG
AGAATAGGAGAAACACTAATCCCATCTAATTCTGCCTTCAAACTCCTAAAATATTCATCA
TTATGAATTAAAAAAAAAAATCAAAGTGTAACCTCACCCAGAGAAAGAAGACATTGGGGC
CAGGTCTGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGATGCTGAGGCGGGTGGATCA
TGAGTTCAAGAGATCGAGACCATCCTGGCCAACATGGTAAAACCCCATCTCTACTAAAAA
ACAAACAACAAAAAAATTAGCTGGGCTTGGTGGCATGCGCCTGTAGTCCCAGCTACTTGG
GAAGCTGAGGCAGGAGAATCACTTCAACCCGGGAGACGGAGGTTGCAGTGAGCCAAGATG
AAGCCACTGCACTCTGGCCTGGTGACAGAGTGAGACTCCGTCTCAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAGGAAAACGAAAAGAAAAGAAAGCAGATATTGGTAATTCT
AGCAGATCCTGGAACAACTGAACCAAATTTATTAATATGTATTATTACTGAAAATCAGTA
ATGAACAAAATTTACAGAATGGGCTTCTTGGAGTTGTTACATTTCCCTTATTACATAACT
CTTCAATAAAAGTGTTTGTCATACCTATTTAGTTAATTCTACAACAACTAGTGTGATAG
GGCTATTATTTGATCTTTTTTTTTTTTTTTTTTTTTACAGGTAGTGACATTCAGTATTA
GACAGCTGCTATTGTGTTAGTTGTCTGAATACCTTTACATATTATCAACTGGCCTTTTCA
TTCCTGAGTTGTGAGTAAATGCTCTGTCTCCCAGACTGGAGTGCAGTGGCGCAATCTCGC
CTCAGTGCAAGCTCCGCCTCCCGGGTTCACACCATTCTCCAGCCTCAGCCTCCCGAGTAG
CTGGGACTACAGGCGCACGCCACCATGCCCGGCTAATTTTTTTTTCTTGTATTTTTAGTA
GAGACAGGTTTTCACCATGTTAGCCAGGATGGTCTTGATCTCCTGACCTCGTGATCCACC
CGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCACACCCGGCCATAAAT
GCAGTCTTGTGTTCCCCACTTCCATTCCTCCTTTGACAGTACAGCTATGCTAGTCTGCGT
AGCAAATTGAAAAAATATGACCTGTGGGATTTAAACAAAACACAGTGTCATACACATTTT
CTGGTAAACTTAACCAAAAGGGACTTGGGTTCCATAACTAATCACCAATGCCTCAGTGAT
CTGTAACTCCTTGTAGGTACCTGATCACAGTTACTAAAGGGAAAGAGGAGCGAGGAATAC
AAGAGCAAAGTCAAGCCAGACATAGATTTTATCTCTTTGTAAACAGGAGTTCAGAAGACC
GCTCTGAATGCTGAGTTAGCATCAGCAATAATAGAAATATATGCAGATTGTTGATTTGAA
GTCATGCAAAGATATCTTTTTCATCCAAATGGAGGCAAAAGCATCATAGAGCACCAGAGG
GCTAAATCCAACTGTAGCAGCAAAAGGTACACAGAAAAATAAAGCATCCTGAACCAACGC
ACTGACTTTCTAGGGCTTATCTAATTTGGAGCTATTTCCTTTTCTTATTTCATTCAGCAA
ATATTTATTGAACACCCACAATGTGTAATCTGTTCTATTACATTCTGTGGAGGAAATACA
GAAGTGAATGAGGCATGGTTCTTACCTACAAGGAATTTCTAATCTTGTGGGGGAGACTAA
CATGTAAACAATAAACTATAGTATGAGGATTACTGAAGAGGCATATGCTAAGTCTCAGAA
CATTGAAATATAAGAGTTGGGTTTGACATGGGGAAAGAAATACCTTCTTCACTGAGGAGG
TAGCATTTTGAGTTATTGTTGACATGTGAATACGATTTTGAAAAGTTCCAAAGAATGAAA
AATTCCACCTACATTGGTGAAGTACTAAGATTAAATGCATGATAGCTTGAAGACACAAAA
ATAATTATTTATAAACCATTCCAAAAATCATTCAGGGAATTCCAATAATACACAAGTTTT
TAAACACATTTCTGGGTAATTTTGAGTAATAAGGTCTTAATCTCCTCTACTGCTTTCAAT
TGTTTTTGTGGCCTTCTTTATTTTGTGGGTATCTGGCCCAGTCTTGTCTGTAGTGTATTA
TGGTGGATTGGATTAAACATGTTTTGCAATCTCTGGAGTGATTTTAAAATGACTTGTGTT
ATATCAGAGTTTCCTAAAGGGAGATTAATTTGGCTTAATGGTAAGAACGGATTAAAGTTA
TGAGATACCAGACACTGGGAAAACAGTTAGAAGCCTGTTGAGACTCTTCAGGGCAGTTGT
TGTGAGAATGAAGTTAAGACAATGGGATAGAATATGAAAAAAAATGAAACAAACATGAGA
GGCAGTCTGAAGATGGAAGTTGGCAACTCATCAAATGTGAGAAATTTATAGGAACAGAAA
AGAACCTGCTGATTAATATAAATTTTCTGCCAAAGAAAGTACAGTGGCTCTCCTCAGCAA
ACTAACATGGGAACATAAAACTAAACACTGCATGTTCTCACTTATAAGCAGAGCTGAACA
ATGGGAACACATGGACACAGGGAGTGGGACATCACACACTGGGGCCTGTTGTCGGGACTA
TGGGAGGGAGACCATCAGGATAAATAGCTAAAGCATGTGGGACTTAATACCTAGGTGATG
GGTTGATAGGTGTAGCAAACTATGATGACACACGTTTACCTATGTAACAAACCTGCACGT
CCTGCACATGTATCCCAGAACTTAAAATAAAATTAAATTATTAAAAAAAGAAAAAGACAG
TGCTTGTCTTATTCGTTTTTTTCTTAAAATGGGAAATATGTAATATATATCAACTGTAGT
GTATAGAAGGGTCATGATGAATTGGACAAAGATACGTGGAGTTTGAATTGCTAGAGGAGT
ACCCACGTGCAGTTGTCCAGCAGAAATCAGGGCTTGTTCCCCAACATGCTATTCACAATC
AGTCTACTACTCTCAGGTATTTGTTTTTCTGTGTGGCTATGCAAGCAATAGATACAGTTT

FIG. 1L

ATGTGAAAATGTTTTAGAAAATGTCTTCTGGAATAATTAAAAGCATACAAGGGAATGTAA
ATCTCTTAATGTGACAAGACCTTTTTGCCACAATAAACAAATTCATTAGTTCAAAAAATA
TTTATTGTGTGCCTATTGCAGCAAACAAAACAGACGAAGCTCCTTCTTGTAGGGAACTTA
TACTCTAGTGATATTTAGTATATATTTTGACAATTGAACCAACAGGATTTGCTGACGGAT
TGCCTTATGGGTATAAGAGAAAGAGAGGAGTCCACACTTTCATGCCAGGTAGGTTGATGG
AGGTGCCATTTACTGAGATACAGGGCCGTAGAGGAGGAGTGTGTTTGCAGCAGGGAAGGA
GAAGACTCAAAATTTGGTTTTGATCATACTAAATTTGATATAGTACAGGTAAGTGTATGG
TGGCCATTAGAACATGAAGGTAAGAGTTTAGATAAGGAGACAGGTATGGTGAAATACATC
CAATATTTATAACCAATATTATCTTTTGTGTCTGTACCTTTTTATACATTCCCCATATAT
ATCAAAGACTATAGAAGGGACTGGATAGTGAATAAGTGATTATACATAAATTCTTTTTTA
CAGATTATTTTGCTCTTGATTTCTCCTATGTAAATCATCACAGCTACATTTTTTAAAATC
TTAAAAAGGATTACTTTGAACAATGCATTTAAACATCCAGAAAACAAAAACAGGAGTGCA
TGGTAAAAATTCTGATTTCAGAACGTATGCCTGACTTATCAAGTCAGAATTTCAGGGAGT
GAAGACCCTGGAATCTACACTTTAAATAGAGCCTCAGTTCACCAAGTATGAGAAGTCCTG
TAACAGGGAAAAGTAACCTCCTGTTATATTTGATGGAGGCCAATTGACAAGCCAAGTAGT
TTTCCATTTGACAAAAATTCTATTGTACCAATGAAGAGCTATCAGAGGGGAGTAGATTAA
AACACCTCCCTTGAAATGGAATTTGGCAAGAAAGCAAGAAATTACAGCAAAAAGACCAAT
AAGAGGAATTAGGGGCAATGAAGGAAGGAGCAAAGATGTGGGAACCCAAAAAGTTTTCCT
AGTAACAACTTTGAAATTATATTTTTAGTATATTAAATTTAAAGTAGAGTTATTAGTGCA
TACATTGGTGTAATTTATTATTATATTAAGCCAACAATATACTTTTAAACTTATACAACT
TTGCAAAAAAGTACAAATCAGAAGTCTGGGCTAAGTAGAATGCATAATAGAATCAGTAGT
GCAAAATATTGTTCTATATTTTCTAGCTTATGATTTTCTATATAAAGTCAGTCTTTCAGG
ATTAAAATGAATGTCACTTCTTTTTACCATGTGTCCTTTAAATTATTAAAATCTATACAC
ATATTGCTATACATAGTAAATATAGTTAGTCAATTATGTCATGGAAAGAATTGAAGGGTT
GTTATAAATTTAAAGGTGTTTCACTATACAAAAACATTGTGAAATACTGGTGCTGATTTA
GTTCTAGTATCTCTGATATATTAAATCATAAATGTCAGGAGTTATTGGTCACAAAATAAA
CACCAGAATTATATGACAGTCTAAAAACAAAAACAAAAAACTTCAGCAACAATATTGAAG
ATATGGAAGTGCCAGAAGAATAAGGATTAAGACAATGAATAAAAATCTCTTCCAAGGACT
GGTCTACACTAAGAGTTTAGAAATGCATTTTTTTTTCACAGAAATATCCTTAATCCTCTA
TATAGAAATGAGAAGAAAACATAAGACTTTAGCAAGCTCCATCTAATCCATTTGCAGACA
TATGGTTACCTATCTTTTCTTCAATATATTGGAGTTTGCAAATATTCTACCTTCAAAGAA
TAGGTGTTACCAAAACATTGTCTGCAAGATTTCTAAGATTTGAAATATATTTGCTATAGT
AGGTTAGAGATGAGACATTTTTACTTTAAATTGCAATAATTCAGACTTAAAATATAAAAT
GTGTAAGTCTAAATTTTTTTTCTATTCATTGCAAATATATCTTATATATACATAAAATCC
TGTGTATACTCATATGAACTTTAAGGAAATATCAGAGGCATCAGTAATAGATAACTTGCA
TCTCTTTTACATTCAGTTCAAGCTACTCAAATTTTAATCTTTTGTTTTCATTCCAACAAA
AAAAATTAGGATCTGCCTTGGCTTTTGCTAAGAAAGTAATTATTGGCTGGACATGGTGGC
TCACATCTGTAATCCCAGTACTTTGGGAAGCTGAGGTGGACAGATTGCTTGAGCTCAGGA
GTTCAAGACTATCCTGGGTAACATGGTGAGAACCTTTCTCTAAACACACACACACGCGCA
CGCGCGCACACACACACACACACACACACACACACACAAATTAGCTGGGAATGATTACACGC
CTGTGGTCCCAGATACTTGGGAGGCTGAGGTGGGAAAATCACCTGAGCCCAGGAAGTCGA
GGCTACAGTGAGCCGTGATTCCACCACTGCACTGCAGCCTGGGTGACAAAAAGAAAGTCA
TTATCTTCAACACTGTGCATACACACTTTTCTGCATCTAGATCCCAAATTTTTGTTTTGT
ATTTACATAGAACATTGATAAGTAAGGTAAGTATTAATTGATAAAACATTTCAAACTCAT
TTTTCACTAAATCCAATGGCCTTCCTCTTTTGCATGAAGTCTCTAAGAATCATGTTAATC
TACATACTCAATCTACGTAACAACTGGATATATCCTGTAGTTGTTGCCCATTTTTCTGCT
AAATGTTATCTTTAGCACTAAGCATGAGTATGAGGAAACAGTATCTGTGCTCAGATTCCA
GAAATGAAGAAAATGTACTGGAGGTCTTTTGGATAATGGCTACAAGGTCACAGGGACTGA
CTCCTTTGGAAGCTCAGCGATAACCATTTTCAGAGAGAATATGTCAACATCTTTCAGTCT
AGAACTTGATGTTCTGCTGAGATCTAATCTGGGGGTGTCCTACTATTGAATAGGTATAAA
CTAAATAAAAATAGTGAGAGAACATTCATGTGTTCACTCATTCATTCCTTCATCAAACAA
ATATTGAAAGTCTATTAATTGGCAAGCACTCTTCTGACATTAGAAGGAGCAAAGATAAAA
AAGATATTATCATTAACCTCAAGGACATGACAGCATCATGGGAAGGCCAGAAATGCAATA
TGTTAAAGTAAAACACAGTGTAGTGTTTACTACTAAAGAGATATAAACAGAGTACTGTGG
TCTAAAATCATATATATAACATTTGCTTAATGGATGAGAAGGAAACTTTAACTTCAGGAG
GCAGAGCATTAAGAAAGTGAATGACAGGAGGGTCAAAAGAAAAAGCCGACAGTGTTGCAG
AGGCAGGGCATAAAGGAGCTAAACCTTTGCTACCTTCAGTTTTTATTATCCACAGAACGA

FIG. 1M

CAAAGAAACAACAACAACAACAAAACTTTGGATTTGAGGGTTTTTTGTTTTTCTTTTTTT
TTTTTTTCCTCTCATTCCAAGCATCAAACCTTGGGATTTATTTACCTTCTAGCAAACCAA
AATTTATGGGGGCATTCCTATGGTCCTCACCTCACCCCATTTTTCTGTTTTACCTATGAA
ACTTGATCAAAATACTGTCTCCACATTTCTCATAAAATACATTAGTTTAATTTTCTACTA
TTACTTTCTTTTAGTTGATTTAAAAAAAGGTCATTTATGACCTATTTAGGTTAGCATCAT
TAATTTTATCAATGTAAGAATATGGTAGTACAGTGTGAATTCCATTAATGGATATGTTGA
TACCATGGGTTTCTCTGACCTTTCCTCTTCCGCTCCTCCCTGATGATTGGTTCTGAGCTT
ATTATCATGTCAGCAATGAAACAGAAAAGGGAGAAAAATCTCAAGTAGGTTGTCTGTCTC
TTTAACACTGAATAAAGATTTTTTTTTCTCTAACAGACTTAAAAATAGTGCCCTAAAAAT
GTTTTGTTTCATTTGTCTGAATTCCCATTCTTTCCCGTGATCATAGATAGTTGAGCTAAA
AAAAGAAAAAACAAAAAACAAAAATAAACATTGTGTCCTACATTTGTATTAACTTTCTTA
GGAATGAGAAGTAGAATCTTAAAAACCTTAGAATGGGAGTTTCCAAGCTAGCTTGCAGGC
TTGAGTTTTATTGATAATACCTTTAGGATGCATGTATTATTAGAAACATCAGTTATTTAC
AAGTTCACCTATTTAAAAGTCTAATAGGAAAAAATATTTCATGTTGCTAAGTATGTGACT
TCCCTTTAAAAGATAATAATGCTTTCCCTTTAAACAACAATAGTAAAAGAAGTAGAGTTC
CTTTTAAACACATACTTTTATATTATAACCCATTCTGTTTAAAAAATAGCAGGCATATAA
TCTAGAAATGCAAATAATTTAGTGAAATTTTTAAAATTATTCTACATATAATTAAATATG
GATATTCGTTTTCAAATATCAAATAATAAAATATGTCTGAGATGCTGACTAATCCTTAAT
TATAGGTGTGATTTCTACTTCACCATCAATACTATGGTACTCCAAATCTTAACATGAGTC
TGATTTTCTAATAAACATGATGAAAAAAGTTATGGAAAAATTTTGAGATTTACTTTGGGA
GGTTCTATTGTGTTCTGTTCAGCTTCATAATATTCAGTTTCTATGAGTTTGGTATTTAAT
TATGTGTGTTTGTCATTCAGTAGGCTGGAAGTATGACCATTGGGAGATCAAAACGATAAG
ACATTAATGACAGTGCTTTATCACTGAATCTAGTACTTTTTTTAATGAAAGAGATGTTGG
CCTCTTGTATTGTTATAAAACAACACAATTTTATGGCTTTAAATTAAAGTACAATCATAA
CAGAAGACAAAATTAGATTAAAAAACAAACATGGAGTGACTCATATAAAATATTTAGAAA
CCAATAATACAGATAGAGACACATTAGTTCCTCTAGACATTGTGTTTTCCAGTAAAATGA
TCACCAAACTTACCAGGAAAATGATAATTATCAGATTATTTACTTTCAGAATTAAAGGCA
GGAAGAGAAAAAAATGAATGAAGAGGAAACACAGTAACCATATAGGACAATAAGAGTGAA
TGAAGATAAAATGAAAAATCAATAAGATATCGACTTTCTTAAAAGACAAATATCACAATA
GGAAACACCTCAGAAAGGGAAATCTCAAGAAAATAATAAACTGAAAGAAGAAAACATATC
AAAACAACTTGAGGACTGACAAAGTTTTAAAATGTATTTAGATAAAGATACCATGAGGAA
AGTGATCAAGGTGTTCTAGGTAATCACTGAAGATAAAACTAAAAATAGCTTAAATTAAAA
TCAGATAGAGAGAAGGTAACTGAAACAGGCATAGAAAGAAAGTAAGAAGGAATACAATCC
TGAACATCTTAACAATGTCTCAAATGTCAGGAATTGATCCAGTTTTTGGCTGCACAACAG
AGTGGCTATAGTTAACAATAATTCACTGTATATTTCAAAATAACTCAAAGAGTAGAATCG
GAATGTTGCTAACACAAAGAAATGATAAATTCTTGAGGAAATGGATATCCCAATTACCCT
GATTTGATCTTTACACATTGTATGCTTATATAAAAACAGTATTCATGGCCGGGCGTGGTG
GCTCACACCTGTAATCCCTGCACTTTGGGAGGTCGAGGTGGGCGGATCACAAGATCAGGA
GATTGAGACCATCCTGTGAATGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGC
CGGGTGTGGTGGTGGGCGCCTGTAGTCCCAGCTACTGGGGAGGCTGAGGTGGGAGAATGG
CATGAACCCAGGAGGCAGAGCTTGCTTGCAGTGAGCTGTGATTGCACCACTGCACTCCAG
CCTGGGCGACAGAGCGAGACTTCGTCTCAATAAAACAACAACAACAACAACAACAAAAAC
AAAAACAGTATTCATAATAATTAAAATAAATTATTTTTAAAATAAAATAAAATATCAGTA
ATTTAAATTTTTCCTATAGCATAGAGATCTGTAATTAATACTTGTCGATCATTGTTGTTT
CTGTCTTCCCAACAACTACACTCCTGTTTCTTCACATTCCCCCTTCTTCTAACAGCACTA
CATCTTTCTTTAGGAAACTATCCTTTTGCCATTTCATGTATATGGTGGGGTGGGGGAGTT
ATCAATCACAGTACCCCAGCAGATGGGACCAGAGGCAAAAATGCCTGACCTTCTCCCATC
CCCCAACCACAGCAGCAAATGAATTATAATTTGATGCACAAGGAAGTATCGGAGCTTTTG
TGTTGGGTTTTACATATCACCTGTGGGAGATAAATGAACTTTTCCCCACCTAACCTTTAG
CCACTTGGGATGATTAGACATAGAGGTGCCTAAGATCTTTCCCTTTGCCACATTAAAAAC
AAATCATCTATGGCACGAGCATACAAGACCAGCTTTCAGAGACACAAAATGATGGAGAGA
ACCATGATACTAGTTTTAGACCTAGTCACTGAGACTTTCTCTGCTCCTTCCCAGTTACCT
GAGCTTTATTTTGTTTACATTTATCAGATTTGAATGGCTGTACTTCAAAGTACTGATTAA
AATAGGAACCAACCTATATGATTCAGGTGGTGAGAAGGAAGAAAAAGAGAGAAAATGAGG
TTAACAAAAGAGAATAAAGAAAAAGAAAGAAGGAAAACAAGAAACTCTGACTACCTCTCC
TCTTTGACATAGTTTACACTTCTGACAGATTGTTCTTCTCTAAATTTATGTAGAGATTAG
AGTGAGGATGATGTATGCACTGTAGCATGGGTGGTCTTCCAGGAAGCCTTGACTGAATGA

FIG. 1N

GGCAAGGAGTATGTTGCTCCCTCAGTAACCTCAAATTTACCTGCAAGCCTGATAAAAATC
TAACACTAACACTAAACCCAATCTTATCTACAGCCCTAACTGCACCCTAATATTAACAAC
CCTACCTCTGTACTTCAAAACTAAAACTAATTCTGATTTTACTCCCATCTGCCCCTTTTA
CCCTAAAACCAACTGTAAAACTAAATTTAACTCTAAACGTAATCCTAAAACTAAGAATTA
ACTAACAATTTTATCTCTATACCCAACTGTTAACCCCAAGCCTAACTCTAATCCTATCTC
TAACCTAACATTAACCACAAACCTACTTCTAACTCTAATCCTAACCATAACCTCAAATCT
AACTCTGATTCCAATTGTAATCTAAACACCAACCCACCCCTACCCTTTTATCCCAAATCC
ATCTGAAACCCTCATCAGAACACAAATTCCAATTCTACTTCCCACCCTGACTCTGACTCT
AAACATAGGCCCAAATATAACTCTAACTCGAAGTCAAAAACTTAACAAACCTTATCTTGA
AACTCAACCTTTACACTAACCCCAATTCTATCTGTAATCCTAACCCTAATATTATCATCA
AACCTATGTCTAACACGACCTCCAACCCAAAACCAAAACTAACCTCAGACCTAACTCTAC
ATCTAATTATAACCCAAACCCCAGGCTGCTACTTACCATAACCCTGAAACTAAGCTTGAT
CCTTTCTCTTTTTTTTGAGATGGAGTCTCGCTCTGTCTCCCAGGTTGAAGTGCAGTGGCG
TGATCTCGGCTCACTGCAAGCTCTGCCTCTCAGGTTCATGCCATTCTCCTGCCTCAGCCT
CCCGAGTAGCTGGGACTACAGGTGCCCGCCACCATGCCTGGCTAATTTGTTGTATTTTTG
GCAGAGATGGGGTTTCACCCTGTTAGCAAGGATGGTCTCAATCCCCTGACCTTGTGATCT
GCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCACGCCCAGCCGAC
CCTTTCTCTTAACCTACACTAACACTAACTGTAAACCTAGCTGTAACTCTAATTGTAAAC
CTAACCTGATTACTCACTACAAAGGTCCCTCTAATTCTAACAAGAAACTCAATCCTATCT
CAATTCCACCCAACCCCAAAAGTAAATCTAAACTTAAACATAACTCAAAATGTATCTCAA
ACCTTAACCTTCACGAAACTACATGTAGCACTAATGTAACCCTAAGCCCAATCCTATCAG
TAACACTAATGCTAAAACAAACCCCAATCTGTATCTCTACCCCATCACTGACACTACCCA
AATCCCAATATTTAATGCTAATCATTAAGTCTCAAACTAACTCCAAACTTATCTTTAACT
ATATCTCCAACCCTAACCCTAACATTAACCCCAAACTTATCATTAATCATACATCTAGCT
CTAAACCTAACCCCAACTTTAACTCTTACCCTAGCTCTAAAATTAACCCCAACACTATCT
CAAACTGTAATCCTAATGCTAACGTTCAGGCTACTTTTAACCCTCACCCTACACAAAATC
CTGCACCTAAACTCAACCCTAACTTTAAACCTACCTCTAATCCAAACACTAAATTTAAAT
CTGGATCACGACTTGGGCATACTAGCACCCACTAGTGTTCTGGGTGTCATTTCTTTGCTT
CACTCTCATCCAGCTTTCTTTACTAATTTTGGATAATGAATCAGAAAATAGTGGTGTGGA
ATCAGGGTCTTTGAATTATTGTATTATCCAGAGTTTGTCCTGCTGCAAATGATAAAACAC
TGAAAATTAGCTTAACAAGGAAAAAAATAAAAATGTGTGTGCAGGGGAGAAGTAGGAGAG
TTATTGGCTCAAGGAACAGAAGAATTCAGTACTGAGTTTCACAAATACCTGGATTCCCTA
GTCCCCATACTGCCATCAGGATCCAATCTGTCACTCACTCCAATCTCTTCCTCCTCCCTG
TTGGATTCATTGTTAGGTTTCCTGTGGCAAGATGAAATGGCCTCAGGCCTGTAACACAAT
AGGATCAATTACAACAGAAGATAGTATTTCTGTTTTCCTGGTTGCTCAAGCCTAAATTCC
AAGATTAGTTTATATCAAACCTAGTTAGTTTTGCTCATGTAAGGGATTACTGCAACTGGG
TACACTAATATGAAGAGTGGGAGAGTTGGTTAAGGGGGTTCTCTGAAAGGAGAATTAGGT
TACTGTTAATGGGAGAATGAGAAATGGGTATTGTAATGACAAAAACACACACGACTACCA
CAAATGTTGAGGAAGAATTTTCCTTTATGATCATCTAGCCCAACTTTTTAATTTCTAATT
TTGTGGTTTTGACCAGTTTTTTGTTTTTTTTTTAATGCAGCATGTCATAAAGTTGGGAA
TACTTCACATTTTGTCTTTGAAAATTTGGAGAGTACTTAAAAAGATTTACAAAGGGGAGG
ATTGAATTATTTTAGGAATGTAAATGGTGGTCTTCTGTCTCAGGCAATGTAGATGCTTGC
TAGAAAACAGCTGACTCATGACTGTTTTCTTTCTAATTCATTAATATGAATTATTTCAAA
CTGCAAAGTTATCTCCTTTCTTCTCCTAATCTATCCACTTAGAGTATACATGTTCAAATT
AATGTATTGAACTAATTTTTCTAGTAATACATTCTATGCATTACAAAAATAGCAGTGGGA
AGGTGAAAACAAAATGCAGTTATGCATTTATCTCTAAATGTGTTCAACATCTCTTATGCG
TACTTCAAAATAATTACATTTGTTTAATTTTGAAAAAAATATTAACAAGAAGTTGTAATT
TGGGGAAAATTTAAAGCTGGCGAAAAAGGCTTCATCATAATTGACAATATGGGAAAATAC
TGTATTAAAATCCTAGGTTTCTCCCTTGTTTGCATGAAGGAAATGAAAAATATATAAGGG
AAGGATTTAATCAGTCAGGCAAAAATCTAAATTCATCACAGGTTTATTCACTGCATACTA
TCAATGTGCCCAGTACCTGAATGAATATATTAAAGAAATCCACCTCTTGTACAATGAATG
TAAATGAGCAGAGTGTGGTGATTAAAGGTTGGTATTTGGTGCTGGGTAGACCCAGCTTTG
CCACTTACTGCCCAAGTAAATATTGCCATCCATCAGATATCTCCACCTATCAGACCCACC
CTGTTGTAATAACAAGATTAAAATCTGTATCACTAAAACTTTAAAAGAATTTATAGCCGA
ATCTAGAAATCTTTCACTATAATTTATCTTTCCTTAAAATGTCGTTTTTTTTTCAATTTT
ACTATATATGTTTTTGTACTTGTTTGCTGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTG
TGTGTGTGTTTTTGAAACGGAGTCTTGCTCTGTCGCCTAGGCTGGAGTGCAATGGTGGGA

FIG. 1O

TCTTGGCTCACTGCAACTCCCGCCTCCTGGGTTCACGTGATTCTCCTGCCTCAGCCTCCT
GAGCAACTGGGAGTACAGGCGCACACCACCACACCCAGCTAATTTTTGTATTTTTAGTA
GAGATGGGGTTTCACTATGTTGGCCAGACTGGTGTTTTTGAAAAGACTTTTTCCTGATT
CAGAAGGTGGGACTCACAATTGTAATTCTGCTAATGGTTGTCTTTCAGTCTATCAATTGC
TTCATAAATGCATCCACTGTTCCTTCTTCTTCTGCCCTGCTTATAATTTCCATGAGTCC
ATATATCTTTTTACACTGTCTTTAGTCTTATTCACTAAATTAAAACTAATTTTTGATAT
TTGGTATTCATGACAAGACAATTAGTAGAATTTTGATGCTTCTTGTCTGCAATTACAGAA
TCAATATATTTTCTATATTATTGTATATTCTCTAAATCTTATTTGTATAATAGCTTTCA
GCATGTTCTTTAATTCTGTTTAGATATTTAGAAAGTATTTGTTGTTATTCTGTAATTTAT
TTCAATATTCAATTATAGTTTAATATTTTGTTATCTAGTGTTGTCTTGATTTTGATATAC
GTACTGATTTTGTAGATCCAAATTCCTCTTTCCTATCAGAGAATGCAATTTTTTACTTGG
ATAAATAAGAATCATATCTCCTCTGCTTGCTACCGTATTGCATACATTCATGGGTAGAGA
AAGAGTTAAGCTGATGAGAGTAGGAATTAAGGTAGACCTGTTTGGTAGGTTCTCCCAGAT
TTCAGAGGACAGACATCTTTTTTCCCTGCCTTGGTCATTTAAACTTTTTGGATTTTGGA
TTAAGTGTAGGCAGGGAAAATGTATCAGATATTTTTATTTTTCTTTGGTGCCATTTGTCC
TTCTCTGCTTTAGGCAGAGAAGCATATGTAGTCCAAGAATGTGCTTTTCTATCCAGCTAC
ATCAATAATAACAATTAGTAAAATTCTACTTAAACTTAGACCTTTGCTGTTCTCTTTTCT
CTGCTTGTGTTAAGTCATGCTCATGATTCTGGCAGTTTTCCACAGTACCATGTACAGAAA
GCTTGAATAAGGTACATCTAGAATACTCATATATGTTCACTTCAAAAACACATTTTTGTG
GAATTCTAAATGCAAATCTCAATAGTGCAATTCTAATTTACAATGAGAAAAAACTAAGGG
ATTTTTTCTGGTGATTCTTTTTGCTCATTTATAAATATGTTTTTAAATGGTAAGCAAATA
TATAAATTAAGCTTTTCCTTACGTAGCTACATTGATTTACTAGTGGTGGAAAAGGTTAAG
CAAAACTAATTTTCATGAGTGTAAATGAATTAGTAAGTGACATATGCAATGCTTAAGGGG
AATTTGCATAAATCTATGACTGATACTCAACCTCTTGCTTAGCGAGAAGATAATTAAAAT
ATTTTATACTTCAAGAAGACCTAGTTTTCCAAATTATTTACATCCACAAACTCAGATTTT
ATAGCAAGTAAGAAAAGTTAAGTCAGAAGCATATACTATTAACAGCTACTTACATTGCTC
AAATTTAATATACGATTGCTGCTTTTGTTGGTTTTGAAATGTTTCTTGACCATGGATCTG
AATAATGAAGTTATTCAAGAAGCAACTTTAAGAATGTTATATTCTTAGAAAGAAGCTATA
GATACAATAATATTAAAAAATTAAATGTAAGTTCCTGCACTCACAGTAGAGGTAAGTTCA
AGGTTATAAGAGAGCTTATAGATTCTGAGATTTGGAAAGAAGAGAATAGAAAAAACTTTT
CAGATTAAATAATGTGTTAATTGTGCTTCTAAAACAGCTTTGGTGATCTTAATAAAATAA
ATATTGTTTTTATTTCCATTTTTGCTTTTCAGACAAGAAATGCTACTTGATGGCTGCATA
TATTTGTTTTGTCTCTTTTCACCACCTACTCTTGCTAAATACTCTCAACCCACTCATGAA
ATTAAAGCACATTGGAAAACATTTATCAACTACCTGTAAATACAACCTATGCTCTCTTTT
GTGGAGGTGATAGACATTCATCAATGGAATAGTTGATCTAAATCCTAGTCTTCATTATCT
TGTTTTATACATTCTTGTCTTAATCAGTTTGGGCTGCTCTAACACAATACCATAGACTAG
GTGGCTGATGAACAACAGAAATTTGTTTCCGACTGTTTTGGAGACTGGGAAGTCCAAGAT
CGAATTTTATGTCTGGTGAGGGCCTGTTTCCTAATTAATAAACATCTGTTGTCTCATATG
TCCTCACATGATAGAAGGGGCAAAGGAGCTCTCTGATGTCTCTTTTTTAGAATATTAATC
TCGTTCATGAAGGCTCTGCTCTCATGACCTATTCCTTCCCAAAGGGCCCACTTCCAAAGA
CCATCATATTAGGGATTAGGTTTCAACAAATGAAGCCAGGGGGAGGTTGGTAAACATTCA
ATCTATAGCAATGCCTATCTCCAGGAGCTGCCTGTGGAAACACTTTTATCTGATATGGTA
GTTTAAAGCATGGCAGGGATAAGTGGTATGAGGAAAACTCTCCCTGCCACCCAACGCACA
CATCCCACTTAAGCTTCAGCAGCTCCAATTTTATCTGTGTAATATTTGGTTCCACATCAA
AGTTGTTTTGAATATACTTCCATTACCTTAAAAAATGTAAAAACACTGCTTTAAAAAGCC
AAGCCTATTCCCTTTTCATTATTCAGAGTTCTTCCAGTTTTACCGTTACATCAAATTAGA
ACTACATAATTAGGAACCCCTCTCTAAATTTGCCTCTATACAGAGAAAAACTGTGCCTGA
AACTTTATTAAAACTCAATAAAGGAAATATGTATGAATGTATATATATAATTTCTCTGAA
GGACAGAATTTGTACTTCGTTCCATACATAAAAACTCATTTGACAAATAACAAGCATAGC
TCCAAGCTCAAAGAATAGCTTAATTTTTCCTGATTAGTTTATATCTCTCTTATTAATCAA
TGACATTTAATATTACAACCATAGCTTGGGGTTTTAGTTTATTTGCTTTCTATCTTTTTT
ATACTGTCGGCCTACCTGTGCCCAACTATGTTATAGTCAGGGGTTGGTAAAATAAAGACA
AAACAAATCCTGTCTTCCTGGAGATCACCTTCACTGGGGGTTGAGAAACAATAAGAACAA
GTAGTAAGTAAAATATGTACATTAAAATTTTAGATGAAGTTAAGTGCTATGGAAAAAAGT
AAAATGGAAGAGGTGTTATGGAGTACCTGTTCGGGTATGGGTTCAATTTACAAGTGGATG
GTCACCTTCTCACTGATAAGGTGACATTTGAGCAAAAGTCTTCAGCAGGAAGGGAGAATG
CCATGCAGTTATCCTAGGAAAGAACATTTCCAATATAAGTAACAGCCAGTGCAAAAGCCC

FIG. 1P

TGATGTAGATGCATACCTTAGGTATACGAGTAACAGTAAGAAATTAGTGGCACGAAAGAC
AGATGTACTTGGAAACCAAAAAGAATCTCTGGTAAGAAATTGTAAGTCATTGTAAGGACT
TAAGGTTTTTTTTTTTCCTCTCCAAATGAGATGGAGATCCATTAGAAGGGTTTGCGTAGA
GAAATAATATGATCTGACTTATATTTAACAGGACTACTCTTTTGCTGAATTGAAAATTGT
CTCTAAGGGTGTATATCAGATCTTATATTGATCTTACCCTTCTCTGTTCAATATTTAACA
CACAAGCCTGTTAAATAGTCCATTCCCAACTTCTGTGACTTCTTGCTTGAGAGCCTTTCT
ATCCCCTCTCATAAGGGCTGTGAGGGCCTAATCTGCTTACCTATCCAGCAGGCTGGGAAT
GACACAGAGCACTCACCAGGAGCACTCTCAACCTATGACTCATGGAAGTTGGTAGATGAA
TACCCCAGCTCTCATATTCCTTGGGTGGAAGAGCTCTGAGATGTGTGTTCTACACCATTA
CCCAGAGGGCACCCTCTGGATTAGGCTCAAGTTGCTGACAGTAGTATCTTGCTGACTAAC
ATAATTTTTATTAATTTTCTCCCCATTTGACCTTATTTCTCCATTTTTCTAATAGTGTTC
ATTGGTATCACTTCCAAAATAAATTACCTTTACTTGAATATTTTTCTTAGAATCTTCTAT
ACAAACCTGAGCTAATACTGGGGCAAAGAGTGGAAGCAGGGAAATATTTTGTAGGTGTTG
TGGTGATGTAGGACAGAGCCTGATAGCTTGGATCAAGGTGGTAGCAAAGGAGATTGTAGA
AGCTATCACACTCTTTATATATTTTGAAGACACAGCCAAGAGGTTTGGTGGAAAAAATGGA
TTGTGAGAAGTAATAAAAAGAGTGGGAGAGAAAGTCAAGGATGTCACCAAAGTTGTCCTA
AGCAAGTGGAAACTTAGATTTGGGAGAATCAAAAATCCTAAAATATCCAAATCCTCTCCC
CTGCCTTCCCCTCCCCTCCCCTCCCCTTCCCTTTGGAGATAGGGTCTTGCTCTGTTTCAC
AGGCTGTAGTCTAGTTTCGCGATCTCGACTCACTGCAGCTTCGACCCCCTGGGCTGAAGT
AATTCTTCTACTTTAGCCTCCCAGGCAGCTGGGACTACAGGATTGCACTAATGTGCCCAG
CTGATTTTTTTTAGTTTTTTTTATTTTTAGTGGAGATGAGGTCTCGCTATGTTGCCTGAG
CTCAAGCAATCCACCCTCCTCAGACTCCCAAAGTTCTGGGATTACAGGTGTGAAACACTG
TGCCTGGCCCAACATTTTATTTTCAAATATTTAAGTTTTGAATGTCTATTCGATAACCAA
GTAAAGAAGTCAACTAGAATATATGAGAATGGAGTTTTCTAGAGAAGTCTGGGTTGAGGA
TGTACTTTTGGGAAATGGAGCACATACTTGGTATCTAAAGCTGTGAGCCGAGATGAGATC
ACTAGGTAGGTAAATATAGATAAATTAGAGAAAATATCTAATAATTGAGACATGGAGTAC
TATCATAAATTTTGAAAAGACAAGAAAATGTGAGAGATCGAGAAGAATGGCTGGGGAAGA
AGGAATCTAAGGTAGTGAAGAGATTGAAATGTGTCAAGGAGAGAAGAGAGTAATTAGCTC
AAATGCTACTGATAAGTAAAGTGAAATGTAGAATGAAAGTCAACCATAAAATTTGGCATT
ATGGGGATCATTAATGACCTTAAAGAAAGTGCTTTTAGTGTAGTAATAGAAAGATGCAGA
AAGTAAGTAGAGTGAATTCAAATTCAACAGAGAATAGACAGAGAGGAATTGAAGACATTT
ATACTGACAATTCCTTCCAAGACTTCTGCTATTAAAAAAAATAAAAAAAGAAGGAGAAAT
GGCAAGTGTTTGGAGGCCAATTTATACTCAAGAATAATTTCTTGAGTTGGTTTTTTGTGT
TTGTTTGTTTTTGATTGGTTAGTGTGTTTATTTTTTAGACGGGATTGGAGAAATACTTTC
ATTTGTGTTTTTACCCATGTTTTCAGCCTTGCCCTGGCTGCCTGGTATAACGCAACTCTA
TTTGTTATTCTGCTATTATAGTTTCCCTAGCTTGAATTTTTTTACACCCTTATTATAATT
GTAGCGTTGCATGCCTATTTCAAAACATCTCATGTACCCCATAAATATATACATCTACTA
TGTACCCACAAAAATTAGAAATAAAAAAATTTAAAAATTATGATTTTTTAAAATTTGTTA
AATAATGTTTACTGACTCTTTTATTTGTTGAAATCATTCATTTTTTGGAATATCAGGTCC
AATTAAATATTTAATCAGACTTTGAGAAGGATTTAATAAGACCAATAAATAACCAAGTAT
TAGTTGAAGGAAATTTCAGATATTTTGGTAGCAGAAGGAACTGAGTTATGGCTCAAGAGT
TTTTTAATAAGTGTGAGTGGAGTTATACAAACTACTCATTAAAATCTTTATTTGAATTTG
TAATATCTGAAACCATTTTCATATTGAAGAATCACTTAAAATAGTCATAAAATGTAAAAT
TGCAAGACAATTAAAAACAAAAATATGATTTCACGACTGTGATAGTACCTGAGAAATTTC
TTCATCTCCTTAGTAAGAGAAGTATTACACCTATTTATAGTTATTTTATGAAACTAGCTA
AGATGAATTATGTAGAAAAGATACAGATTTTCAAACAGAAACTAGAATTAATGGAAGCTA
TGTGAGACTATAAAGAGTTTAATAGTTATTTTGATTTTTTTTTATGAGTGCAAGGAGTAT
AGCGAAAAATAGCATCTACCTATAAGGATTTGCAAAGCCAGTAATCTTTCTAAAAATATC
AGCAAACCCAGAATTAAGGCTTATGTTCTTAGCTCATTGTAACTAGATCAAAAATAAAGA
AGGCCAAATAAAGGTATGTGACATTTGTTGAAAACCTGAAGTGTCCTATATGCAGAAATA
TTTTTATCATTTAATTAATTTCAGAAACTTCTTAACATGACATGATCCTCTTGAAAAGAT
CACATCAAAAAAGGCAAAATAATTGCATAATTATTGTAGAATAATTTTTGTGTGAGTATT
TTTGACTTAGTGTAAGTTTCCAAGTTCAGGATTTATCATGCAGTGAAAAAAAATACACTT
GTCTAGAAGACAGGAGACTTCATTATATTCCTCTCTTTACAATTAATTAACGTAAGACCA
TTTAAAATATGCCTAATTTTCCAGGCATTGGTTTGCTTTGCTATAAAATGGGAGGATAGA
AAATAACTTTCAAAATATCTTATAAATCTAAGAATCTTTGCATCTTATAAATCTAAGAAT
CTTTGGAAATTCATAGATTATTGAGATGGAGTCTCGTTGCTATGCATTGTAGCAAAGTTG

FIG. 1Q

GAAATAAATTCTAAATTTTATTTCATTTATATTGATCAATAAATTGTTACATTTCACTAA
TACAATAAGGAAAATTTATTTTACCTGAGTGTATGTCTAGCTTGTGAAATAAAAATGCTC
AATTATGAAAGCATTTATTGCCATTTTGAATGAAAAATGTAATATGTAGAACAGAATTTT
TTTTGCCTTGAACTCAGTTAAATGTAGAAATTGATAAGGACTTGCATTTTCATGAACTTA
ATAATTATCTGTCTTTTCAATGGTCTCCATATCAAGTCTGAGAAATATGGATGTGATTTA
TTTTAAACCTCACCATTTGAAGTAAATCTAAAGATTCCATTAGGTTATGAGCATATAGGA
TACAAGGACCATATTGACAGTTTTGTGGGATTGTATTAGGATAAAAGGGTAGGAACAATG
GGGAGAAAATTATAGCTTACAATAGGGAAGAACCAAAAATTGTTGCAAAATGATGGAACA
GGCTGAAAGAATGATATAACCTCCTAAACACTTCAAATGTTTAAGCAGTTCATTGTACCA
GGGCCATTGTAGCAAATATTTTCTGTCTTGGGTGGAAGGTCAGTCAAGGTGACTGATAAA
GTTTCTTCTAACGATAAAATAGCACAACTCACTTTTTTTCTAACCTCTAAGAGTATATTA
ATATCAAAAGAAGGCAAGCAACAAACTACTTCTGAATGTTAATATATATCTGCATTCATT
TTAAAAGTCTGCTACAACTACAGATAGAGGAACAGTTTGTAGTATCCGTGATCCTAGAAC
AAATTTAGCTTTTAATATCTTGTCAACTTTTTTGTTTTAGTATCTCTTCCTTGGAACTAG
CTGAGCTTTAATGGCATCATCATGTGATATGACTTGAGATTTATATTTGGAAGAGCTTTG
AAAAATCACGGATTGTTACCCTAATGAGGTGTTATTCAGTCTTTTAAACAAGAGCAATTT
CTTTACAAAAAGGAGCAGAATTCTTAATTGTATCTGTAAACCTCCATTTAAGAATGAATT
ACTTGGCTGGGCATGGTGGCTCACACCTGTAATCCCAGCACTTCGGGAGGCAGAGGCTGG
TGGATCACTTGAGGTCAGGAGTTTCAGACCAGCCTGGCCCAACACGGTGAAAAACAGTCT
CTACGAAAAATAAAAAAAAAAAAAAAAAAAAAAAATAGCCAGGTGTGGTGGTGTGTGCCT
GTAATGCCAGCTACTCGGGAGGCTGAGGTGAGAGAATCACTTGAACCTGGGAGGTGGAGG
TTGCAGTGAGCCAAGATTACACCATTGCACTCCAGTCTGGGTGACAGAGCGAGACTCCAC
CTCAAAAAATAAAAAATAAAAAAAAAAAGAATGAATTGCTCATAAATGTGCCTCACTGAT
GATTAAATTTAATCCTGCAAGATTATGTCTTTTGATGGAAATGAGAGGGTTTATACAAAG
TTTTATTCGTGATGTTATCTATGTCATCTATTGATTTCTGCTCTGATTCATGTGGATGAA
GTTACACCTCACACTTTAAGCTGGTGTCAGTCTTCCCATTTTCTGCTGTGATGTGTACTC
AAGATCTCCAGATTACATCTGTAATGTAATGCAGCCATGATTGTTTATAGGTACATTTAG
ATGAATTCAATGATGAGTTATGTTGTAATAAGTGTCAGATTTAGATGAACCATACAAATA
AAAGAACCATGCATTAAAATGACAAATGTGTAAAAGCATTATTTGGGCCTTAAGTCAAGG
CCCAAATGTGGATACTGGTACTGAGACATCTTTCAGAAAGGAGGTATGAAGTACTGAAAA
ATATTTACAAAATGAAGACTACTTTTATCTTACTTATCATGATTCTTTTATTACATATGC
ATTTTCTAAGATAACTATAGTGCATTAGTTTGTACTATGTTAATATAATAATAGGGTAAA
TCAAACAATGTTTTCTAAATCCATTAAAATAGAGTTCCCTAAGGGAGTTAAAACAATTAC
GTTCTACTGTATATTATTGGCATGCTTCAGGAGACATGATTTAATCTCTAGACTATCAGA
ATTCAAGAACTAGTGAGTCATATAACAAAGGAGGCTTAATCATGCCATTTAAGTGTCATG
GAAAAAGGTTTATTGGTCAGGAAAAATTAATTAGAAAAAAGTTATAAAATACTTCACTAA
GAAAATAAAATGTCAGGAAGCCCACTTAGACAATGAGTGAAAATGAAACAAATTCAAGTT
TTTACAATATTTGGTTTCTATAGGATTGCTTCATTGTTTTGGTTTTTGTTTTTCCCCATA
AGCTGATCTCAGAAACTTTTCCTCTACATGAAGAGGCTGTCATTTTTTCATGGTGTGTGT
TTGTTCACATGCCACACAGACAATCAATTATGAAGAAAGGAGAGACTCGTAGGAGGCAGG
GCCAGGCTGTTCACACTTTTAAACTAGGTAGCCACAAATGAGGCTTAGTTACAAAAACTT
GAAAACTGGATTCTTCCCAATGTATTATACATCCCCAAAGAAATGATGAAGTTCCTTACT
CTCTTCTCTTTGTTTTTGTAAATCTTACCACTTCAAGTGTTGGCAATACTTACTTTAAAG
TAGGTTTTCATATTGGCTTAGATTTTTTTTTCATTAACTTGCAATTTGTGGTTGGGAAAT
GATCTGCTTTTTGTTTCAGGTTGTTTAATGTTTTCCAATGTAATATTCTTCTTGCACTCC
AGTGAGTTTATTTACAAAACATTTAATGTCATTTGCGTCTTCGAAGAACAATGTATTCGG
TTAGAACAAAAGTGAGCTCCTGCATAGAGCTTATGATGGTTTATAATTGGTAAATTATTA
CCTTGGTCAAGTTTGTAAACTAATAAAGGGAGTAGAAAACTTTAGATAAAAAAAACTAC
CTCATTCAAAGGGACCGTTCACCCACAAAATGCCTTTTTGTTTATCTTTTGGAATGACAC
CATTGGAAACTCAGTATGGCCACTTTTATGGTAATAATAAAAGTCATATATAAAAAGGAT
TATTAGAAATGTGTTATTTCTTAGGCAGGTATGCTTATTTAAAGTATGTATGCATACATA
CTTTAAACTACTAAATACAAATAAATTAGTAGTACAGTCATTAGGATTGCTCTTAGTTTG
TTAGTGTTGGAATAGACTTTTGGATTTTCTTCCTAGCTTAGATTGATACAATGTGATGGG
GACTTGCTCTCCAAACACAGGAATAGGTGGCCTGCAGACACACTCTGTGATGCTGTAATT
CTAATCCTCACTGAATATATCAGGGGTGGACATCTGGCCTGGGGCAATTCAGATACTTTT
TCTTAAAATTTATACTACAAATTCAAAAGTGGTAACTCATCTCTGCCATCACTTATAGTA
GAATAAGACCCACTGTTGCAGTGGGGAATTGAGAAACCCAGTCCACAGGGAGAACAAACA

FIG. 1R

TGGAGAATAAAATAAGTAAATTAGAACAGGAAAAATGCCAAAACACACAGACATGACCCT
GATAGTTTTCCATTTCCTGATCACTGTCCCTTCCTGTGGCTGGATAAGGAACTGTCTCTA
GGCTCTGTAAGACATATTTGCATCCTTACGACAAATTTCTACTCCTTTTCATAAACTAGA
CTTGGGTTCTTTAACTTGCAACAGCAACAACAATAAACGATTTTGTTGGGTACAATCTGA
TTTTATTAACTTCTGGATTTAAAAGCCCTTCTAAATGTTGATTGGCATTGTTTTTACTTC
CTAAGAGTACGCTCATGCACCACATAGTGATGTTTTGGTCAACGACAGACTGCATTTACG
ACTGTGGTCCCATAAGATTATAATACCATGCTTTTCTGTACTTTTCTATGTTTAGATATG
TTCAGATACACAAATGCTTATCATTGTGTTATAATTGCCTACAGTGTTCAGTACAGTTAC
ATGCTGTACAGGTTTATAGCCTAGGAGCAATTGGCTATACCCTATAGCCTAGGTGTGTAG
TAGGCTATACCATTAGATTTGTGTAAGCATACCCTATGATGTTTGCACAATGATGAAATC
ACCTAAGGATGCATTTCTCAGCATATATCCCAGTCATTAAGCAAAGACTGACTCTATTAT
TAGGTCTATTTTATTCTATAGCATTTGATCATGAGATATGTGAAAATAAATATAATTTTT
AGAAGTACAATAACTTTCAAATCCTGAATGTTCTGTACTTTCCATCTCACAAGCATTTTG
CAAAGCATCAAATGGTATAAGCCAGATTACTGTTAAGGCAACTTGGAATTAATATGCTGC
TCAGTTCTGGAAAAGGCATATTCTGTAAATATAGATGAGAGAATATAGACTTTTTCCCTC
TCTTCTTACAATCCACATTCTATTCAGTATTTCATTTACTTGAGGGGTTATATGCTACTT
ATCTTTATCTGTTGTGGAGTGAGGACACATTCCAAATGCCTTGGTATTATTAAAAGCCCT
TCATGATGTGGCCCCATCTTTTATGACTTTTCCTTTTCAACTGTGCCCTCTAGCCTTATT
TGATTTCTCTCAAATTCTTAAACACAGCATGCTTCACTGACCTTTAAGCCTTTGCACATA
CAGTGTTGATGTGGAGCTTCCTGACCAACTCCTAATTCTCCTTCAGGCCTCAATTTAAAC
ATCACTTCCTCTGGGAAGCTTTCTATTATTCCCAAGGTACTGGGATATGTTCTTGCACAG
CATGCTGGGCTAATGTCACAATGGCTACCTTGTTTTATTGTTAGTATTTGATCAGCGACA
CCTTGCCAGGGAGCCCCTGAGTATTGTCTGAGCAGAAACTATGGCTATCTTGTCCCCTGT
TTAGCACAGGGCTTCTCTAAAAGTGGGCTTCTCTAAAAGTAAGTGCTCAAGAACAACAAC
AAAAAGTGTTACATTAATAAACACACACACATACATACAAAGAAATACCTGTCTTTCTCC
ATATCTCAAGATCATGCTGAAAAGCCAGCATTCATGAACAAATTCCTGTGCGAAGATTGA
GAATGAAAGATGAATAAGAGGTATCTTTAGAACCCAATTATGGCTGCCGTTGTTCCCTGA
GTGTGAGGCTTGCTGTTAGAGTGACAGAAGGAATTTTGACTACTCAAGACCATACAAATT
TGGAAATGACTCCAAAGTAAACATGGTTAGATAACTACACATTCCATTCCCCCTTTTTA
TTTCTATAGAATCCCAACTTTGTTCAAGTAGTAACATGCCCAGCTTCAGAAATGAGTCAT
GATTTTTCTAAAGCAACAATATCAATCTTCTTTCCCTTCCCCAGTGATTGGTATGGAAGT
GGACATTTCAGCAAGTTTTAGCCAATAACGTGAATTCTGTTTTGAAGCATCTAAGAAAGA
TTTTGCTTTCTGCTGTAAATCAAAAGCAGAAACAGGAGAAGATTCTTTTGGGCCTCTTTC
CCTCTTCCTGGCGTGGAAGTAGTTGTGAGAGCATATGATACCCAAAGTTTCGGTAGACAT
TTTATAATTATGTGATGAATAACCTAAGGATAATTAAACATATAAAAGAATGGAGAAAGA
CTGAGTCTGTTTTACTCCACAAGATGCTGAACCAACCCTGAGACATAATTTATCTGGATT
CTTAAATAACTAGTGTCTTTGTGGTTTAAGCTGTTCTTTGTAAACAAACATATCATAAGT
GATTAAGTGATGTTATCTTCCTTTAAGGCAATCAAAATGCATCTGACAAATGGCCATCTA
ATTTAAAATTCCAACTATGTAGACATCTCAAACAAAGTCAGTATCTCAAAAAATATACTA
CAAAAATTCTCATGTGTCCATTGGGGATAACTTCCAATGCTCTTTCATTGGTATTGTAGC
TATGGCATTTGATTTCCAATTGTATGTGGATCAGGTAGTTGCAGGGTGACTCTCAAGGGC
GAGAAGAAAGTAAGAGTACATGAAAAAAAAGAGGAAGAGAGAGAGCAGACAAGAAGGAAG
AACAAGACAAAGTCAAACCCTAGGTAGAAATAAGAAGGAGCTAGTACAGAAAGCAAATGC
CTAAGGTGTTGGAGAACATAGAAAGGTAGAGTGGAATGAAAAAGAAAAAAACACTAAATA
GCAGCACATAGAATCTTGGGGTTTCAGGGATATTGTTTATGAAAGGTTAGAATAGGCAAC
AATCTACCTTGTGGCATCTTCTTAAAATTATCAACATATAAAACAAACAATAATTATTTA
AATTACCTGTCATATGGGTCTTGTCATTTATTTATAATTTAAGGAGAATTAAAACTGAAC
TAGTTGCTGGGGAGTGACATCAGCAAGATGGAGATATAGAAATCTTCAGGACCTCCTTCC
GTCCATGGAACCACTGACTCAAAAATGACAAATGGAAAAAATTTACTTTCTGAGAAATCA
AGAAGCCAGTTAAGAGGCTCCTGTATCTCAGATGAGTGCAAAGCCAGCTGCAACAGAGCC
AGCAGAAAATTTGTTGTACTCACTCTTCATGGTCACTTCTGGCATAGCACAGTGCAATCT
AGAAGAAATTCTCGGCTCCTGACTACTTTCTTGGAAAAGAAAGAGAAAAATGTACCATAT
GTCTAATATTCTGATGGGGATGGGGTGTGGGCTGCTCAAAGGACTAGCTTCCGTCATGCC
TAAATACAAGTGCTAATTGGGAAGTCCACAATGTTGGGGGCTGCAGAAAACAAGGGCAAC
AGTTTGGACTAGCATGCACTCATTTGCCGCAGTTCCTCCTCTCACTTCATAGAATGAGTA
GAAGAACCCTTAACTCTCAAGGTTTTTTCCTGGGGAGAGAAAGAGTCAAAGCAATTATA
CAATATTATGGCTTTGTGGGAGTGATGTATCCAAAAAAAAAAAAAAATGAGTTTTTACCAC

FIG. 1S

ACCAATCTCAGAGTGCAGATGGAACCTAGCATATTCTAGATGCCTGGGGGCCATTGAGAA
CAAAAGAGAGCTAGGCAACTTTCAGCAGCTCCAGAAGAACTGTGGTACCACAGATAGACA
CCAAAGGGAGGAAGAGATTACAAGCTCCTGAAAAAAGAAATGAGCAATTCATTCTAATTG
AGAATTTACACACACTGGTACAGATAAGATGAATTTGCAAAAAAAGAATAGAGGCCCCAG
AATTTCTAGCTGGGTTTTTTGGTGAAGGCCTTTCTCTGTATCAAGCTAGTCCCTAAAGAC
TGGGTGAGGTGGTTTTTGTTTGTTTTACATTTTTATTTTAAAAGATGGGGATCTCACTTT
GTCACCCAGACTTGAGTGCAGTGATGCAATCATAACTCACTGCAGCCTCAAACTCCAAGG
GTCAAGTGATCTTTCCACCTCAGCCTCCTGAGTAGCTGAGACTAGAGACACATGCCACTG
TGCTTGATTAATTTTTATTTTTTTATTTTTTTTCGTAGAGATGTGGTCTCACTTTGTTGT
TCAGGCTGGACTTGAACTATTGACTTCAAGGGATCCTCCTGACTCAGCCTCCCAAATCAT
TGGGATTACAGGCATGAGCCACCATGCCTGACCTGTTTTGTTTTGTTTTAAAAAACTCAG
AAAAATTTCAAAATAGCAATTATAAAGACAATGAGCTTAGAAAACCAATTAATGGACAAA
ATGTAACTATAAGTAAAGAGATACATGTAAAAAGAATCAAACAAAATTTGCAGTGGAAGA
ATATGATAACCAAATTGAATATTACATTAGAGGAGTTTAATACTAGATTTGAACAAGCAG
AAGAAAGAATCAGGGAACTTGAAGATGGGTCATTTGTAATTATTCAGTCAGAGAAACAAA
AAGAAGACTAAAAAAGAGTGAAGAAACCCCTAAGGACATCATCAAGTAGACCAATATGTGT
TATCAGAGTTTTAGAAGAAAAAGACAGAAAAATAGGCATAAAGCATCATTGACAAAATAA
TGACCCAAAACCTCCCAATTATGAAAGACAATAGATATTCTGAATCCAGAGCACAATGGC
CTGCAACTAAGATGAACCCAGAAAAGTCTATACTTCAGCACATTATAATCTAATTATCAA
AAGCCAAGGACAAAGAAGGAATTTTGAAAGCAGAAAGAAAATAGTGACTCATCAGATACA
CAAGGGCTGTCATGAGAATATCAGCAGATTTCTCAGCAGAAAACTTGCAAAACAGAAATA
AGTGGGATTACATATTCAAAGAGCTGAAAAAAAGTCTGCCAACAAAAAAATCCTTTATCCA
GAAGAATTTTCTTCAAAATGAAGGAGAATAAAGGATATTCCAGATAAACAAAAGCCAAGG
GAATCCATCACAATTAAACCTGCCTTACAAGAAATGCTAAATGAAGTTGTTCAAGTTGAA
ATAAAAGAACGCTGAACAGCAACACAAAAGCATATAAAAGTATAAAGCTCATTGGTCAAA
GATAGATATAAAGGAAAAACAACGGGATATTATAATGGTGGTGGGTAACTTACTCTTCAT
CCTGGTATAGAAGTTAAAAAAAACCACAAGTATTAAAATAACTGTAACTATAAAATTATT
AATGAATACACAATGTAAAAATATGTAATTTGTGATACTGATAACATACCATGTGTGGAG
GGGAGAAGTCAAAGTGTAGAGTTTTAAATAAGACTGAGGTTAGGTTTTTATCACCTTAAA
ATAGATTGTTATAATATGTTTGATTTAAGCCCCATGGCAACTACAAAGAAAATACCTACA
GGTAATAAACAAAAGAAAATGAGAAAGAAATGAAAGTGTGTCTCAGTCCATTTTTATTTT
GCTATAACTAAACATCTGAGACTAGGTCATTTATAGAGAAAATAAATTTATTTCCTGCAG
TTCTGGAGGCTGTGAAGTTCAAGACTGAGTTGCTGCCTCTGTTGAGGGGCCTTCTTATTG
CATCATAACATGGCAGAAGGCATCACATGACAAAAAAGCAACAGCAAGAGCCAAACTGGC
TTTTATCATAGGCCTAGTTTGTGACACCTTACATAGTCCTATGAAAACCCATTAAGCCAT
TAGCCCATTAATCCATTAATTCATGAATAGATTAATACATCCATGTGGGGAAAGCCCCTCA
TGACTCAAACCTTTCTCAAAAAAACCCATCTCTTAATACTGTTACATTAGTATTAAGTTTT
AACATGAGTTTCAGAGTCTAGAAATATTCACACCATAGCCTTTCACCCATGACCTCCCAT
AATTTATGTCCTTATCATATGCAAATACCTTCATTCCATTCCCGTAGCCCCGAAGTCTTA
ACCTGTTCTAGCACCAACTCTAAAATACGAAGTCAAGAGTCTCATCTGAGACTCAAGGCA
TGATCCATCCTTGGGCAGGTTCCCTTTCAGTTGTGAAATCAAAACAAGTCATATAATTCT
AAAATACAGTGCTGGTACAGGAATAAGACAGACATTCCCTTGTCGAAAGGGAAAATAAAC
TAGAAGAAGGGGTTAATGGTCCCCAAGCAAGTCTTTAACACAGCAGGGCACATATTAAAT
TGTAAAGCTAAAGAATACTCTTTTTTGGGTCCATGTTAAGCATTCTCTGCACAATGTGGG
GAACACATTGAGCCACTCTGCCCCTATGGCTTTGCTGTGCTCAGAACACACTTCAGCTTT
CTCAGATTGGAATTGCTCATTGGTGCCTGCAGCTTTCCCAGGTGGGCACTGCACACTGCT
GGTGTTTCTATAATTCTAGGATCTCAAAGGCAGCTCTGGCTCTCACCCCGTATTTTTACT
CAACATTGCTGTAGTGGGGCTCTCAGCCATGGCTCTGTCCCTGTGACAAGTCTCTGCCTG
GGTCCCCATGCTTTTAGATACATCCTCTGAAGTCTAGGTGAAGGCCATAGTGGCCCTACA
ACTCTTGCATTCTGTATCCCTGCAGAATTAGCACCAGGTGGACACTGCCAAGGCTTATGG
CTTTTGCTTTCTGGAGCAGTGAGGTAAGCTACACTTGGAGCCTCTTGAGCCAGTTGGAGT
GGCTGAGGAATGATGCGCTCACATGAAGGGAGCAGAGGAGTCCTGAGCAGCCCTGGGCAG
CAAGCTGTGGAGAGTACCCTGGGCCTGTCCCCTGAAACTATTCTACCCTCCTTGGCCCCT
GGGCTTTTCATGAGAGGGGGGCAGTCTTAAAAATATGCAAAATACTTTTCAAACATTCTCC
TCATTGTCTTAATGAATAACATCTGACTCCCTTCTATCAGTGCTAATCTCTTTAGCAAGC
AGTTTTGCTGTTTACATGGCTAAGCAAGCTGCAAACTTTTCAAATCATTTTGCTGTGATT
CCCTTTAATTATACATCTGTCTTTAAGTCATGTTTTTGCTCCTGAATTGGCCAAAAGTAA

FIG. 1T

CCACACAGCCAAAAGTAGCCAAACAGCATCATGAATGCTTTGCTCCTTAAAAATTTCTTC
TATAAGATATTTTACTTTATTATTGTCAAGTCTGGCCTTCTACACAGCCCTAGAGTATGG
ACACAGTTCCAGTAAGCTTTTTGCTACTTTATACCAAGTATGACCTTTATTCCAGGTTCT
GATACCTTGTTCCCCCTTTCTGTCTGAAACCTCATAACGGCCTTCATTGTCTATATGTTT
ACTAGTATTTTGGCCATAATCACTTAAATAATTTATAAAATGATTCAGACTTTCCCTAGT
CTTCTCATCCTCTGATCCTTCACCAGAAGCACCCCTTAACACTCTATTTACAGCAATATAA
GATTTTTTTTGCCTGCTCCTCCAAACCCTTCCAGCCTTTGTCCATTACCCATTTCCAAAG
CCACTTGCACATTTTTAGGTTGAGCATCAGCCTCACTTCTTGTTACCAAAGCCTGTATTA
GGGTTCTCCAGAGAGACAAAACCAATGGGATATACAGAAGGGGATTTGTTAGGGAAATTG
GCTCACACAGTTATGGAGACTGAAAAGACCAAGGTCAAGGGGACGTATCTGGTGAGAACC
TTCTCATTGTATCATAACATGGCAGATGGCATCACATGCTAAAAGAGCAAGAACAATAGC
CAAACTGGATTTTATAACAGACCCACTCTTGACGACTATCCTATTCCTGTGATAAGCCAT
TAATCTGTGAATCCATGAGTAAATTAATCTATTCATGAGGGCTCTGCCTCTATTGTCCCT
TAAAGGCCCCACTTCTTAATACTGTTACATTGGGGATGAAGTTTCAATATGGGTTTCAGA
GGAGACAAACATTCAAACCATAGTGATGTCACTACAAAAAAAATTAATGAAACACAAAGGA
GTACAGTAAGAGAGCAAAATACAGATAAAAGTGCTATATGATATATAGAAAACAATAAAA
TGGCAATAGTAGGAGTTTATCTGTCAGTAGTTACTTTAGCCATAAATGAACTAAACTCAA
ACAAAAGACAAAGATTAGCTGACTGGATTTAAAAAATACTATATGCTGTCTACAAGAAGT
ACAAGGAGCCCACTCCAAATTTGTAGACACACATAGGATAAAATTAAAGGATGGAAGAA
AGTATTCCATGTGAATGGTAACCAGATGAGAGCAGGGCTCATTATACTTATATCGGACAA
ATAAATTGTAAGTCAATAATTGTCACAAGGAACAAAGAAGGACAATATGTAATATTAAAA
GAGTCAATTCACCAGAAAGATATAACAATTTTAAACATATATGTATTCAATCTTAGGGCT
TTAAAATATATAAACAAATATTAATGGAACTGAAGGGAGAAAGACAGCAATACAACAATA
GTAGGAGATTTAATTCTCAGCTTTCTTTTTCTAGAGACAGAGTCTCACTCTGTCACTCA
GGCTGGAGGGCAATGGTACAATCTCAGCTCACTGCAATCTCCACTTCCCAGACTCAAGTG
ATTCTCCCACTTCAGCCTGCTGAGTAGCTGGGACTGCAGACATGCAACACCATACCCAGC
TAATTTTTTAACTTTTTGTACAGATGAAGTCTCGTATATTGCCCAGCTGGTCTTAAACTC
TTGGGCTCAAGTGATCCTTCACCTGGGCCTCCCAAAGTGCTGGGATTATAGGCATGAGCC
ACCGTGCTCAGGACCCAACTTTCAAAAATTGATAGAACATCCAGACAGAAGATCAATGAG
AAGCGGATTGAACAACGTAGACCAAATAAGCCTAACAAACATATGCAGAAAATTCCATCT
AACAGCACCAGAATATGCATTCTTCTAATGCACACACACATATTATCCAGAATAGATCAT
ATGCTGTGTCACAAAACATGTTTTAACAAATTTAAAAATACAGAAATCATATCAAATATC
TTTTCTGAACACAGTGGAATGAAACTATAAATCAATTATAAAAGGAAACTGGCAATTTCA
CCAATATGTGTACATTAAACAATAAATTCTTGAACAGTCCATGAGTCAAAGAAGAAATTA
TAAGGGATATTTGAAATGTTTCAAGATAAATGAAAATGTCTCAAGATGAAATAAAAAGAC
AACATATCCAAATTTATGGAATGCAACAAAAGTGGCAAGAGTTAAGTTTATAGTGGTAAG
TGACTACATTATAAAAGAAAAAAGATTTTAAGTAAACAACCTAACTTTACACCTCAGAAG
TGGAAGAAGGAGAAAATACTAAGCCTAATGTTAGCAAAGAAAGGAAATAATAAAAATTAG
AAAAAATAAATTAAATAGAAAGTAGAAAATTACTATAATAATTAATGAAACTAACAGCTG
CTTTTTAAAGATCAATAAAATTTACAAACCTTTGGCTAGAATAACTAAGAAAAAAGAGAG
AAGACTCATAAATAATATTGTAAATAAAAAAGGAGCTATTGCAATCAAAGAGGCAGGAAC
AATAAAGATTTTCAGGCTATTCTGTATAATTATACACTAACAAATTGGATAACCTAGAAG
AAATGTATAAATTCTCAGAAATACACAACCTACCAAGACTGAATCAAGAAGAAATACAGA
ATCTGAACAGATCTGTAACTAGTAAGGAGATTAAATCAATGATCAGAAACTTCCCAAAAA
AGAAAATCCCAGGATCAGAAAACTTCACTGGAGAATTCTGCCAACATTTAATAGAAAAAA
AAATGCCAATTCTTCTCAAACTTTTGCAAAAAAATTGAAGAGGACGAAGCATTTCAAACTC
ATTTTATGAGTCCAGCATTTTCCTGATACCAAAATGAGATAAAGATATTACAACGAACAC
ACACACTTTCAAACAAGCTACAGGCCACTATCTCTGATGAATGTAAATGCAAAAGTTGTC
AATAAAAAATAGCAAACTGAATTCAACAGTGCATTAAAAGGATCACACACTGTGACCAAG
TTGAATTTATCTCTGGAATGATGAATGGTTTAACATATGAATATCAATCAATGTGATACA
CTATATTAACAGAACAAGGGATAAGATCACATGATAATCTCTATAAATGCTGAACAATCA
TTTGACAAAGTTTAATACCCTTTCGTAATAAAAATACTCAACAAACTATGAATAGAAGGC
ATGTACCTCAACACAATAATAAAGGTCACATATCAAAAGCTAACAGATAACATCATACTC
AATGGTAAAAACTGAAAGCTTTTCCTCCAAGATCAGGAACTAGGTAAGAATGTCCATTCT
TGCCATTTCTCATCAACGTATTACTAGAAGTCTTTGCTAGAACAATTATGCAAGAATAAG
AAATAAAAAGCACTGAAATCAGCAAGGAAGAGGGAAAATTATCTCTATTCCCAGATATAA
TAATCTTATATGTAGAAAATTCTAAAAATCACACAAGGAAACTGTTGCAACTAGTAAGTT

FIG. 1U

CATCAAAATTGCAGAACATAAAATCGAAATGCAAAAATCAGTTATGTTTCTATACAATAG
CAGCAAACTCTCTGAAAAAGACATTACAATCCCACTTACAATATTATCAAAAATGACTAA
AATGTTTAGTAATAAGCTTAACCAAGGAGGCTAACGACTTATACACTGAAAACCATAAAA
GCATTACCAAAAAATAATTTTAAAAGACACAAATAAATAGAAAGATAATTCTGTTTTCAT
GGGTTAGAAAACTCGATATTGTTAAAATGTGCACACTGCTGAAAGCAATTTATAGATCCT
ATACAATCTTACCAAAATTATGATGTCATTTTTTTCAGAAATAGAAAAAAAAATCTGAGAA
CCATGGATACTTAGAAAATCTGGAGAAAGAAGAGCAAAGTAGAGGGTCTCATGCTTCCTG
ACTTCAAAACATATTCCAAAGCCATTGTAATAGAAACAGTTTAGCACTGGCATAAAGACA
GATATATGAACTTACAAACCAGCATAGCGAGCCCAGAAATAAGCCCACACATACATTGTA
AAATAATATACAAAGCACAAAGACTATGGACAGGATAGTCTCTTCAACAATTGTGTTGGG
AAAACTAGATAGCCATATTCAAAGGACTGAAATTAGACCCTACTCAAAAAATCAAGTCAA
AATGAATTAAAAATTAAAGATCTGGGCCGGGCGTGGTGGCTCACGCCTGTAATCCCAGCA
CTTTGGGAGGCCAAGGGGGTCAGATCACGAGGTCAGGAGATCGAGACCATCCTGGCTAAC
ACAGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGCGTGGTGGTGGGCGC
CTGTAGTCCCAACTACTCAGGAGGCTGAGGCAGGAGAATGGCGTGAACCTCAGAGGCAGA
GCTTGCAGTGAGGTGAGATCACGCCACTGCACTCCAGCCTGGGGGACAGAGCAAGACTCC
ATCTCAAAAAAAAAAAAAAAAATACAAGATCTGAAACTATGAAACTCATAGAGAAAAACAG
GAGAAAAGTTTTATACCATTGGTTTTGGCAATAATTTCTTGTATACGACACCAAAGAACA
GGCAGTAAAAGCAACAAAAAATAGATAAGTGGAACTACATAAAATTAAAAACTGATGCAC
AGAAAATAAATAAAAAGAAAAAACAGAGTGTAAAAGCAAACCATGAAATGGGAGAGAATA
TTTGCAAACCATATATCTGATAATGGGTTAGTATTCAAAATATATAAGGAACACCTACAA
CTCAATAGCAAAAAACTAACCCAATTAAAAATGGACAATGGACCTGATGGATATCTCTCC
AAAGAAGATGTAAAAACAGCCAACAGATACATGAAGAGTGCTTAACATCATTAGTAATTA
GGGAAATGCAAACCAAACCACATGAGCTATCATCTTACACCTGGTAGGATGACCATTATG
AAACAAAAGAAAGAGAATTAAAAAAAAAAAAGTGTTGAAAGGGATGTGGAGAAACTAGAA
CCTTTGTACAGCCACTGTGAAAAAATGTTTGGAAGTTCCTCAAAAAAATTAAAAATAAAA
CTATACGATCCAGTAATCCCACTTTTAGATACTTTTCCAAAATATTTGAAAACAGGAACT
CAAAGAGATATTTGCACTCTCATGTTTATTGTAGCCTTATTTACAATAGTCAAGAGGTGG
AAACAAATGAAATATATAATGACAGATGAGTCAATAAAATGTGGCATGTACATATCATGG
AATATTATTCAGCATTACAAAAGAAGAAAATCTTATAATATGCTGCAACATAGACAAACC
TTGAGGACCTTATACTAAATAAAATAAACCAGTCACAGAATGACAAATACTGCATGAATA
TACTTCTATGAAGTATCTAAAGTAGTCAGTCATAGAAGCAGGAAGCAGAACGGCAGCTGC
CAGGTCCTGGGAGTAAGAGTAAGAGGAAAGTTGCATTTCAGTGGGTATAGAGTTTAAAGC
ATGCAAGATGAAAAAGCTCTAAAGATCTGATGTACAATAATATGCATATAATGAACAATA
TTGTACTGTTCACTTAAATATGTGTTAGGTCCATGTTATGTGATTTTTACCACATTTTTT
TGAAAGCAAGTTGCTAAAGAATTTGCCAAATGGAATTATAGTGACACGAGTTCAAATAAA
ATTAAAAAACGAGAAACAGTAGAGTTTACTTAATTTGTTAATATATCCATATTATCATTT
TAGGGAATTTTTACTAAAGCAGAGTATATAAACTATCTTTTTTTGTTCTAATGATCCATT
TGTTTTAGTTTGTTTCCCATTTTTATGTAGCTAGACTGCCAGTTAATCTCCTAAAATTAT
TGGCACCATATTTCCCATTTTTTCTGGCTTTTTTATTAGTAACTGGGATCCTTGCAGCTG
TATCTATGTGATGCCAAACAATTAGGTTGATCAATTCTGTGACAACAAGCCATCTGGTTA
CTTTAGTGAATAGGCCCCTTACTTACCTTTCATAAGTTGATTCTATTCTCCTTTGTGCCTT
CTCTTTAAATTACCATTATCCTGTAACCATAAATTAAAAATACAGCATCGCTTTTAAAAC
ATCCTGAAGTAATTTTTAACACTACAAAAGAGAAGAAATTTCCTTTGTTTGGTGTTCTTT
GACCCTAATTAGCATTTAGGAACAAACTACACTTGCAAAATTATTTTCGATTGGTAGAGG
GAAGAAAAGGGTCTTTTTATTACTATGTATTTGTAATTACTTTTGTCACTTATGTTATTC
TTGTGTCTAAATTCAACTCTAGATTTATTCTCTGTTGATATTTTTTATCACTTGAGAATA
TTTTAGTTTTTCAACCTCTATATGGCGGGCTATCACTCCAAATTTAGGTTAAACTGTAGG
TTGATTTAAAAATCTGGCTATGATGCAGAAAAAATTCGGGCAACTTACCTAGAAAAAAAAA
AGTAGTTATATTTCAGTACTTCTTTTACCTAATCAGCCATTTTAAAATAATTTTGTTCAT
TATCAATATGGAGGAAATTATTTATATGCAGGGAAGTTATTTATATGCAGAGCTGTTAAT
GGCAGCAATCTGCATGACAAATTTCTACTTAATAAGCAATGAAATAGTTGGATAAATGTG
TATTTCTACATGGGTGAATTTCCCAAAATTCACACTTCAAAGACAGTTGCTGACATTTTT
TCAATGAGAGATTTTATTAGATAATGAGTCATCTTAGAGTTATCTTGTAAGTATTCTTTA
GTCTTAATTTAAATTTAAATGAAAGTCAATTCAAAGTGTTGTATTTTCTTAAATAAATTT
TGTTTTATAAACATTAGAAATTAAATAGGACTACCATATGGTCTAGCAATCACACTTCTG
GGTATATATCCAAAGAAAATCAGTTCAGTATGTCAAAGAGATGTTTCGTATTCATTGCAG

FIG. 1V

AQ

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 May 2001 (03.05.2001)

PCT

(10) International Publication Number
WO 01/30991 A3

(51) International Patent Classification[7]: C07K 14/705, C12Q 1/68

(21) International Application Number: PCT/US00/23021

(22) International Filing Date: 22 August 2000 (22.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/426,290 25 October 1999 (25.10.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/426,290 (CIP)
Filed on 25 October 1999 (25.10.1999)

(71) Applicant *(for all designated States except US)*: **DECODE GENETICS EHF.** [IS/IS]: Lynghals 1, IS-110 Reykjavik (IS).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **OLAFSDOTTIR, Berglind, Ran** [IS/IS]: Eskihlid 15, IS-105 Reykjavik (IS). **GULCHER, Jeffrey** [US/US]: Unit M, 130 South Canal Street, Chicago, IL 60606 (US).

(74) Agents: **CARROLL, Alice, O.** et al.: Hamilton, Brook, Smith & Reynolds, P.C., 530 Virginia Road, P.O. Box 9133, Concord, MA 01742-9133 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report: 6 December 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/30991 A3

(54) Title: HUMAN NARCOLEPSY GENE

(57) Abstract: The gene for hypocretin (orexin) receptor 2 (HCRTR2), which is associated with narcolepsy, is disclosed. Also described are methods of diagnosis of narcolepsy, pharmaceutical compositions comprising nucleic acids comprising the HCRTR2 gene, as well as methods of therapy of narcolepsy.

INTERNATIONAL SEARCH REPORT

Internat. Application No
PCT/US 00/23021

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07K14/705 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, CHEM ABS Data, WPI Data, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 96 34877 A (HUMAN GENOME SCIENCES INC.; LI YI (US); ROSEN CRAIG A (US); SOPPET) 7 November 1996 (1996-11-07) the whole document ---	1-7
Y	LIN LING ET AL: "The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene" CELL,CELL PRESS, CAMBRIDGE, NA,US, vol. 98, no. 3, 6 August 1999 (1999-08-06), pages 365-376, XP002153571 ISSN: 0092-8674 abstract; figure 6 ---	1-7
	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 March 2001	19.04.01

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hardon, E

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Internat. Application No
PCT/US 00/23021

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	SAKURAI T ET AL: "Oxerins and oxerin receptors: A family of hypothalamic neuropeptides and G Protein-coupled receptors that regulate feeding behaviour" CELL,CELL PRESS, CAMBRIDGE, NA,US, vol. 92, 20 February 1998 (1998-02-20), pages 573-585, XP002105412 ISSN: 0092-8674 page 585, column 2; figure 2	1-7
Y	ALDRICH, MICHAEL S. ET AL: "Narcolepsy and the hypocretin receptor 2 gene" NEURON (1999), 23(4), 625-626 , 1999, XP000973742 the whole document	1-7
Y	SIEGEL, JEROME M.: "Narcolepsy: A key role for hypocretins (orexins)" CELL (CAMBRIDGE, MASS.) (1999), 98(4), 409-412 , 20 August 1999 (1999-08-20), XP000941943 the whole document	1-7
A	LECEA L ET AL: "The hypocretins: hypothalamus-specific peptides with neuroexcitatory activity" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA,NATIONAL ACADEMY OF SCIENCE. WASHINGTON,US, vol. 95, January 1998 (1998-01), pages 322-327, XP002105411 ISSN: 0027-8424 the whole document	1-7
T	PEYRON CHRISTELLE ET AL: "A mutation in a case of early onset narcolepsy and a generalized absence of hypocretin peptides in human narcoleptic brains" NATURE MEDICINE,NATURE PUBLISHING, CO,US, vol. 6, no. 9, September 2000 (2000-09), pages 991-997, XP002153570 ISSN: 1078-8956	1-7

1

International application No.
PCT/US 00/23021

INTERNATIONAL SEARCH REPORT

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 7
   because they relate to subject matter not required to be searched by this Authority, namely:

   Rule 39.1(iv) PCT - Method for treatment of the human or animal body by therapy

2. [ ] Claims Nos.:
   because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
   because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Internat. Application No
PCT/US 00/23021

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9634877 A	07-11-1996	CA 2220036 A	07-11-1996
		AU 715286 B	20-01-2000
		AU 2470795 A	21-11-1996
		EP 0828751 A	18-03-1998
		JP 11505110 T	18-05-1999

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 May 2001 (03.05.2001)

PCT

(10) International Publication Number
WO 01/30991 A2

(51) International Patent Classification[7]: C12N 15/00

(21) International Application Number: PCT/US00/23021

(22) International Filing Date: 22 August 2000 (22.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/426,290 25 October 1999 (25.10.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/426,290 (CIP)
Filed on 25 October 1999 (25.10.1999)

(71) Applicant *(for all designated States except US)*: DECODE GENETICS EHF. [IS/IS]; Lynghals 1, IS-110 Reykjavik (IS).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: OLAFSDOTTIR, Berglind, Ran [IS/IS]; Eskihlid 15, IS-105 Reykjavik (IS). GULCHER, Jeffrey [US/US]; Unit M, 130 South Canal Street, Chicago, IL 60606 (US).

(74) Agents: CARROLL, Alice, O. et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/30991 A2

(54) Title: HUMAN NARCOLEPSY GENE

(57) Abstract: The gene for hypocretin (orexin) receptor 2 (HCRTR2), which is associated with narcolepsy, is disclosed. Also described are methods of diagnosis of narcolepsy, pharmaceutical compositions comprising nucleic acids comprising the HCRTR2 gene, as well as methods of therapy of narcolepsy.

HUMAN NARCOLEPSY GENE

RELATED APPLICATION

This application is a Continuation-in-Part of U.S. Serial No. 09/426,290, filed October 25, 1999, the entire teachings of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

Narcolepsy, a disorder which affects approximately 1 in 2,000 individuals, is characterized by daytime sleepiness, sleep fragmentation, and symptoms of abnormal rapid eye movement (REM) sleep that include cataplexy (loss of muscle tone), sleep paralysis, and hypnagogic hallucinations (Aldrich, M.S., *Neurology 42*:34-43 (1992); Siegel, J.M., *Cell 98*:409-412 (1999)). In humans, susceptibility to narcolepsy has been associated with a specific human leukocyte antigen (HLA) alleles, including DQB1*0602 (Mignot, E., *Neurology 50*:S16-22 (1998); Kadotani, H. *et al., Genome Res. 8*:427-434 (1998); Faraco, J. *et al., J. Hered. 90*:129-132 (1999)); however, attempts to verify narcolepsy as an autoimmune disorder have failed (Mignot, E. *et al., Adv. Neuroimmunol. 5*:23-37 (1995); Mignot, E., *Curr. Opin. Pulm. Med. 2*:482-487 (1996)). In a canine model of narcolepsy, the disorder is transmitted as an autosomal recessive trait, *canarc-1* (Foutz, A.S. *et al., Sleep 1*:413-421 91979); Baker, T.L. and Dement, W.C., Brain Mechanisms of Sleep (D.J. McGinty *et al., eds.*, New York: Raven Press, pp. 199-233 (1985)). The possibility of linkage between *canarc-1* and the canine major histocompatibility complex has been excluded (Mignot, E. *et al., Proc. Natl. Acad. Sci. USA 88*:3475-3478 (1991)).

A mutation in the hypocretin (orexin) receptor 2 gene in canines has been identified in narcolepsy (Lin, L. *et al., Cell 98*:365-376 (1999)); Hypocrexins/orexins (orexin-A and -B) are neuropeptides associated with regulation of food consumption (de Lecea, L., *et al., Proc. natl. Acad. Sci. USA 95*:322-327 (1998); Sakurai, T. *et al., Cell 92*:573-585 (1998)) as well as other possible functions (Peyron, C. *et al., J. Neurosci. 18*:9996-10015 (1998)). Human cDNA of receptors for orexins have been cloned (Sakurai, T. *et al., Cell 92*:573-585 (1998)), however, full human genes for the orexin receptors have not yet been identified.

Diagnosis of narcolepsy is difficult, as it is necessary to distinguish narcolepsy from other conditions such as chronic fatigue syndrome or other sleep disorders (Ambrogetti, A. and Olson, L.C., *Med. J. Aust. 160*:426-429 (1994); Aldrich, M.S., *Neurology 50*:S2-7 (1998)). Methods of diagnosing narcolepsy based on specific criteria would facilitate identification of the disease, reduce the time and expense associated with diagnosis, and expedite commencement of treatment.

SUMMARY OF THE INVENTION

As described herein, a full gene for the human hypocretin (orexin) receptor 2 (HCRTR2) has been identified. The sequence of the HCRTR2 gene as described herein is shown in Figure 1 (SEQ ID NO: 1). Accordingly, this invention pertains to an isolated nucleic acid molecule containing the HCRTR2 gene. The invention also relates to DNA constructs comprising the nucleic acid molecules described herein operatively linked to a regulatory sequence, and to recombinant host cells, such as bacterial cells, fungal cells, plant cells, insect cells and mammalian cells, comprising the nucleic acid molecules described herein operatively linked to a regulatory sequence. The invention also pertains to methods of diagnosing narcolepsy in an individual. The methods include detecting the presence of a mutation in the HCRTR2 gene. The invention additionally pertains to pharmaceutical compositions comprising the HCRTR2 nucleic acids of the invention. The invention further pertains to methods of treating narcolepsy, by administering HCRTR2 nucleic acids

of the invention or compositions comprising the HCRTR2 nucleic acids. The methods of the invention allow the accurate diagnosis of narcolepsy and reduce the need for time-consuming and expensive sleep laboratory assessments.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1A to Fig. 1AY depict the sequence of the human orexin receptor 2 gene (SEQ ID NO:1) and the encoded receptor (SEQ ID NO:2).

The foregoing and other objects, features and advantages of the invention will be apparent from the following more particular description of preferred embodiments of the invention, as illustrated in the accompanying drawings

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to a human hypocretin (orexin) receptor 2 (HCRTR2) gene, and the relationship of the gene to narcolepsy. As described herein, Applicants have isolated the HCRTR2 gene. The gene and its products are implicated in the pathogenesis of narcolepsy, as mutations in a closely related receptor, hypocretin (orexin) receptor 2, have been associated with the presence of narcolepsy in a well-established canine model of narcolepsy (Lin, L. *et al., Cell 98*:365-376 (1999)).

NUCLEIC ACIDS OF THE INVENTION

Accordingly, the invention pertains to an isolated nucleic acid molecule containing the human HCRTR2 gene. The term, "HCRTR2 gene," refers to an isolated genomic nucleic acid molecule that encodes the human hypocretin (orexin) receptor 2. As used herein, the term, “genomic nucleic acid molecule” indicates that the nucleic acid molecule contains introns and exons as are found in genomic DNA (i.e., not cDNA). The nucleic acid molecules can be double-stranded or single-stranded; single stranded nucleic acid molecules can be either the coding (sense) strand or the non-coding (antisense) strand. The nucleic acid molecule can additionally contain a marker sequence, for example, a nucleotide sequence which encodes a polypeptide, to assist in isolation or purification of the polypeptide. Such

sequences include, but are not limited to, those which encode a glutathione-S-transferase (GST) fusion protein and those which encode a hemagglutinin A (HA) peptide marker from influenza. In a preferred embodiment, the nucleic acid molecule has the sequence shown in the Figure (SEQ ID NO:1).

As used herein, an "isolated" or "substantially pure" gene or nucleic acid molecule is intended to mean a gene which is not flanked by nucleotide sequences which normally (in nature) flank the gene (as in other genomic sequences). Thus, an isolated gene can include a gene which is synthesized chemically or by recombinant means. Thus, recombinant DNA contained in a vector are included in the definition of "isolated" as used herein. Also, isolated nucleotide sequences include recombinant DNA molecules in heterologous host cells, as well as partially or substantially purified DNA molecules in solution. Such isolated nucleotide sequences are useful in the manufacture of the encoded protein, as probes for isolating homologous sequences (e.g., from other mammalian species), for gene mapping (e.g., by *in situ* hybridization with chromosomes), or for detecting expression of the HCRTR2 gene in tissue (e.g., human tissue), such as by Northern blot analysis.

The present invention also encompasses variations of the nucleic acid sequences of the invention. Such variations can be naturally-occurring, such as in the case of allelic variation, or non-naturally-occurring, such as those induced by various mutagens and mutagenic processes. Intended variations include, but are not limited to, addition, deletion and substitution of one or more nucleotides which can result in conservative or non-conservative amino acid changes, including additions and deletions. Preferably, the nucleotide or amino acid variations are silent or conserved; that is, they do not alter the characteristics or activity of the hypocretin (orexin) receptor 2.

Other alterations of the nucleic acid molecules of the invention can include, for example, labeling, methylation, internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, carbamates), charged linkages (e.g., phosphorothioates, phosphorodithioates), pendent moieties (e.g., polypeptides), intercalators (e.g., acridine, psoralen),

chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids). Also included are synthetic molecules that mimic nucleic acid molecules in the ability to bind to a designated sequences via hydrogen bonding and other chemical interactions. Such molecules include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

The invention also relates to fragments of the isolated nucleic acid molecules described herein. The term "fragment" is intended to encompass a portion of a nucleic acid sequence described herein which is from at least about 25 contiguous nucleotides to at least about 50 contiguous nucleotides or longer in length. One or more introns can also be present. Such fragments are useful as probes, e.g., for diagnostic methods, as described below and also as primers or probes. Particularly preferred primers and probes selectively hybridize to a nucleic acid molecule containing the HCRTR2 gene described herein.

The invention also pertains to nucleic acid molecules which hybridize under high stringency hybridization conditions, such as for selective hybridization, to a nucleotide sequence described herein (e.g., nucleic acid molecules which specifically hybridize to a nucleic acid containing the HCRTR2 gene described herein). Hybridization probes are oligonucleotides which bind in a base-specific manner to a complementary strand of nucleic acid. Suitable probes include polypeptide nucleic acids, as described in (Nielsen *et al.*, *Science* 254, 1497-1500 (1991)).

Such nucleic acid molecules can be detected and/or isolated by specific hybridization (e.g., under high stringency conditions). "Stringency conditions" for hybridization is a term of art which refers to the incubation and wash conditions, e.g., conditions of temperature and buffer concentration, which permit hybridization of a particular nucleic acid to a second nucleic acid; the first nucleic acid may be perfectly (i.e., 100%) complementary to the second, or the first and second may share some degree of complementarity which is less than perfect (e.g., 60%, 75%, 85%, 95%). For example, certain high stringency conditions can be used which distinguish perfectly complementary nucleic acids from those of less complementarity.

"High stringency conditions", "moderate stringency conditions" and "low stringency conditions" for nucleic acid hybridizations are explained on pages 2.10.1-2.10.16 and pages 6.3.1-6 in *Current Protocols in Molecular Biology* (Ausubel, F.M. *et al.*, "*Current Protocols in Molecular Biology*", John Wiley & Sons, (1998)) the teachings of which are hereby incorporated by reference. The exact conditions which determine the stringency of hybridization depend not only on ionic strength (e.g., 0.2XSSC, 0.1XSSC), temperature (e.g., room temperature, 42°C, 68°C) and the concentration of destabilizing agents such as formamide or denaturing agents such as SDS, but also on factors such as the length of the nucleic acid sequence, base composition, percent mismatch between hybridizing sequences and the frequency of occurrence of subsets of that sequence within other non-identical sequences. Thus, high, moderate or low stringency conditions can be determined empirically. By varying hybridization conditions from a level of stringency at which no hybridization occurs to a level at which hybridization is first observed, conditions which will allow a given sequence to hybridize (e.g., selectively) with the most similar sequences in the sample can be determined.

Exemplary conditions are described in Krause, M.H. and S.A. Aaronson, *Methods in Enzymology, 200*:546-556 (1991). Also, in, Ausubel, *et al.*, "*Current Protocols in Molecular Biology*", John Wiley & Sons, (1998), which describes the determination of washing conditions for moderate or low stringency conditions. Washing is the step in which conditions are usually set so as to determine a minimum level of complementarity of the hybrids. Generally, starting from the lowest temperature at which only homologous hybridization occurs, each °C by which the final wash temperature is reduced (holding SSC concentration constant) allows an increase by 1% in the maximum extent of mismatching among the sequences that hybridize. Generally, doubling the concentration of SSC results in an increase in T_m of ~17°C. Using these guidelines, the washing temperature can be determined empirically for high, moderate or low stringency, depending on the level of mismatch sought.

For example, a low stringency wash can comprise washing in a solution containing 0.2XSSC/0.1% SDS for 10 min at room temperature; a moderate

stringency wash can comprise washing in a prewarmed solution (42°C) solution containing 0.2XSSC/0.1% SDS for 15 min at 42°C; and a high stringency wash can comprise washing in prewarmed (68°C) solution containing 0.1XSSC/0.1%SDS for 15 min at 68°C. Furthermore, washes can be performed repeatedly or sequentially to obtain a desired result as known in the art. Equivalent conditions can be determined by varying one or more of the parameters given as an example, as known in the art, while maintaining a similar degree of identity or similarity between the target nucleic acid molecule and the primer or probe used.

Hybridizable nucleic acid molecules are useful as probes and primers, e.g., for diagnostic applications, as described below. As used herein, the term "primer" refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis under appropriate conditions (*e.g.*, in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer, but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template, but must be sufficiently complementary to hybridize with a template. The term "primer site" refers to the area of the target DNA to which a primer hybridizes. The term "primer pair" refers to a set of primers including a 5' (upstream) primer that hybridizes with the 5' end of the DNA sequence to be amplified and a 3' (downstream) primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

The invention also pertains to nucleotide sequences which have a substantial identity with the nucleotide sequences described herein; particularly preferred are nucleotide sequences which have at least about 70%, and more preferably at least about 80% identity, and even more preferably at least about 90% identity, with nucleotide sequences described herein. Particularly preferred in this instance are nucleotide sequences encoding hypocretin (orexin) receptor 2.

To determine the percent identity of two nucleotide sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first nucleotide sequence). The nucleotides at corresponding nucleotide positions are then compared. When a position in the first sequence is occupied by the same nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical positions/total # of positions x 100).

The determination of percent identity between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin *et al.* (*Proc. Natl. Acad. Sci. USA*, *90*:5873-5877 (1993)). Such an algorithm is incorporated into the NBLAST program which can be used to identify sequences having the desired identity to nucleotide sequences of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (*Nucleic Acids Res, 25*:3389-3402 (1997)). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., NBLAST) can be used. See http://www.ncbi.nlm.nih.gov. In one embodiment, parameters for sequence comparison can be set at W=12. Parameters can also be varied (e.g., W=5 or W=20). The value "W" determines how many continuous nucleotides must be identical for the program to identify two sequences as containing regions of identity.

The invention also provides expression vectors containing a nucleic acid comprising the HCRTR2 gene, operatively linked to at least one regulatory sequence. Many such vectors are commercially available, and other suitable vectors can be readily prepared by the skilled artisan. "Operatively linked" is intended to mean that the nucleic acid sequence is linked to a regulatory sequence in a manner which allows expression of the nucleic acid sequence. Regulatory sequences are art-recognized and are selected to produce a hypocretin (orexin) receptor 2. Accordingly, the term "regulatory sequence" includes promoters, enhancers, and

other expression control elements such as those described in Goeddel, *Gene Expression Technology: Methods in Enzymology 185*, Academic Press, San Diego, CA (1990). For example, the native regulatory sequences or regulatory sequences native to the transformed host cell can be employed. It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the receptor desired to be expressed. For instance, the gene of the present invention can be expressed by ligating the gene into a vector suitable for expression in either prokaryotic cells, eukaryotic cells or both (see, for example, Broach, *et al.*, *Experimental Manipulation of Gene Expression*, ed. M. Inouye (Academic Press, 1983) p. 83; *Molecular Cloning: A Laboratory Manual*, 2nd Ed., ed. Sambrook *et al.* (Cold Spring Harbor Laboratory Press, 1989) Chapters 16 and 17). Typically, expression constructs will contain one or more selectable markers, including, but not limited to, the gene that encodes dihydrofolate reductase and the genes that confer resistance to neomycin, tetracycline, ampicillin, chloramphenicol, kanamycin and streptomycin resistance. Vectors can also include, for example, an autonomously replicating sequence (ARS), expression control sequences, ribosome-binding sites, RNA splice sites, polyadenylation sites, transcriptional terminator sequences, secretion signals and mRNA stabilizing sequences.

Prokaryotic and eukaryotic host cells transformed by the described vectors are also provided by this invention. For instance, cells which can be transformed with the vectors of the present invention include, but are not limited to, bacterial cells such as *E. coli* (e.g., *E. coli* K12 strains), *Streptomyces*, *Pseudomonas*, *Serratia marcescens* and *Salmonella typhimurium*, insect cells (baculovirus), including *Drosophila*, fungal cells, such as yeast cells, plant cells and mammalian cells, such as thymocytes, Chinese hamster ovary cells (CHO), and COS cells. The host cells can be transformed by the described vectors by various methods (e.g., electroporation, transfection using calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances; microprojectile bombardment; lipofection, infection where the vector is an infectious agent such as a retroviral genome, and other methods), depending on the type of cellular host.

The nucleic acid molecules of the present invention can be produced, for example, by replication in a suitable host cell, as described above. Alternatively, the nucleic acid molecules can also be produced by chemical synthesis.

The nucleotide sequences of the nucleic acid molecules described herein (e.g., a nucleic acid molecule comprising SEQ ID NO:1) can be amplified by methods known in the art. For example, this can be accomplished by e.g., PCR. *See generally PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, *et al.*, Academic Press, San Diego, CA, 1990); Mattila *et al., Nucleic Acids Res.* 19, 4967 (1991); Eckert *et al., PCR Methods and Applications* 1, 17 (1991); *PCR* (eds. McPherson *et al.*, IRL Press, Oxford); and U.S. Patent 4,683,202.

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4, 560 (1989), Landegren *et al., Science* 241, 1077 (1988), transcription amplification (Kwoh *et al., Proc. Natl. Acad. Sci. USA* 86, 1173 (1989)), and self-sustained sequence replication (Guatelli *et al., Proc. Nat. Acad. Sci. USA,* 87, 1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

The amplified DNA can be radiolabeled and used as a probe for screening a library or other suitable vector to identify homologous nucleotide sequences. Corresponding clones can be isolated, DNA can be obtained following *in vivo* excision, and the cloned insert can be sequenced in either or both orientations by art recognized methods, to identify the correct reading frame encoding a protein of the appropriate molecular weight. For example, the direct analysis of the nucleotide sequence of homologous nucleic acid molecules of the present invention can be accomplished using either the dideoxy chain termination method or the Maxam - Gilbert method (see Sambrook *et al., Molecular Cloning, A Laboratory Manual* (2nd Ed., CSHP, New York 1989); Zyskind *et al., Recombinant DNA Laboratory*

Manual, (Acad. Press, 1988)). Using these or similar methods, the protein(s) and the DNA encoding the protein can be isolated, sequenced and further characterized.

METHODS OF DIAGNOSIS

The nucleic acids and the proteins described above can be used to detect, in an individual, a mutation in the HCRTR2 gene that is associated with narcolepsy. In one embodiment of the invention, diagnosis of narcolepsy is made by detecting a mutation in the HCRTR2 gene. The mutation can be the insertion or deletion of a single nucleotide, or of more than one nucleotide, resulting in a frame shift mutation; the change of at least one nucleotide, resulting in a change in the encoded amino acid; the change of at least one nucleotide, resulting in the generation of a premature stop codon; the deletion of several nucleotides, resulting in a deletion of one or more amino acids encoded by the nucleotides; the insertion of one or several nucleotides, such as by unequal recombination or gene conversion, resulting in an interruption of the coding sequence of the gene; duplication of all or a part of the gene; transposition of all or a part of the gene; or rearrangement of all or a part of the gene. More than one such mutation may be present in a single gene. Such sequence changes cause a mutation in the receptor encoded by the HCRTR2 gene. For example, if the mutation is a frame shift mutation, the frame shift can result in a change in the encoded amino acids, and/or can result in the generation of a premature stop codon, causing generation of a truncated receptor. Alternatively, a mutation associated with narcolepsy can be a synonymous mutation in one or more nucleotides (i.e., a mutation that does not result in a change in the receptor encoded by the HCRTR2 gene, such as a mutation in an intron or an untranslated portion of the gene). Such a polymorphism may alter splicing sites, affect the stability or transport of mRNA, or otherwise affect the transcription or translation of the gene. A HCRTR2 gene that has any of the mutations described above is referred to herein as a "mutant gene." It is likely that a mutation in the HCRTR2 gene is associated with narcolepsy in humans because of the association between a mutation in the HCRTR2 gene and narcolepsy in dogs (Lin, L. *et al., Cell 98*:365-376 (1999), the entire teachings of which are incorporated herein by reference). In a preferred

embodiment, the mutation in the HCRTR2 gene is to a deletion mutation, for example, a deletion that corresponds to the deletions found in the hypocretin (orexin) receptor 2 in narcoleptic dogs as described by Lin *et al.*, *supra* (e.g., a deletion of one or more exons, such as a deletion of the fourth exon, that can be caused by insertion of one or more nucleotides upstream of the splice site of the exon, or a deletion of exon 6, that can be caused by a G to A transition in the splice junction consensus sequence). In another preferred embodiment, the mutation in the HCRTR2 gene is mutation that effects a "knockout" of the entire gene, such as deletion of the first exon as described by Chemelli, R.M. *et al*, (*Cell 98*:437-451 (1999), the entire teachings of which are incorporated herein). In a third preferred embodiment, the mutation in the HCRTR2 gene is a mutation in an intron, that affects splicing (joining of exons) during translation of the HCRTR2 gene.

In a first method of diagnosing narcolepsy, hybridization methods, such as Southern analysis, are used (see Current Protocols in Molecular Biology, Ausubel, F. *et al.*, eds., John Wiley & Sons, including all supplements through 1999). For example, a test sample of genomic DNA, RNA, or cDNA, is obtained from an individual suspected of having (or carrying a defect for) narcolepsy (the "test individual"). The individual can be an adult, child, or fetus. The test sample can be from any source which contains genomic DNA, such as a blood sample, sample of amniotic fluid, sample of cerebrospinal fluid, or tissue sample from skin, muscle, placenta, gastrointestinal tract or other organs. A test sample of DNA from fetal cells or tissue can be obtained by appropriate methods, such as by amniocentesis or chorionic villus sampling. The DNA, RNA, or cDNA sample is then examined to determine whether a mutation in the HCRTR2 gene is present. The presence of the mutation can be indicated by hybridization of the gene in the test sample to a nucleic acid probe. A "nucleic acid probe", as used herein, can be a DNA probe or an RNA probe; the nucleic acid probe contains at least one mutation in the HCRTR2 gene. The probe can be one of the nucleic acid molecules described above (e.g., the gene, a vector comprising the gene, etc.)

To diagnose narcolepsy by hybridization, a hybridization sample is formed by contacting the test sample containing a HCRTR2 gene, with at least one nucleic

acid probe. The hybridization sample is maintained under conditions which are sufficient to allow specific hybridization of the nucleic acid probe to the HCRTR2 gene. "Specific hybridization", as used herein, indicates exact hybridization (e.g., with no mismatches). Specific hybridization can be performed under high stringency conditions or moderate stringency conditions, for example, as described above. In a particularly preferred embodiment, the hybridization conditions for specific hybridization are high stringency.

Specific hybridization, if present, is then detected using standard methods. If specific hybridization occurs between the nucleic acid probe and the HCRTR2 gene in the test sample, then the HCRTR2 gene has the mutation that is present in the nucleic acid probe. More than one nucleic acid probe can also be used concurrently in this method. Specific hybridization of any one of the nucleic acid probes is indicative of a mutation in the HCRTR2 gene, and is therefore diagnostic for narcolepsy.

In another hybridization method, Northern analysis (see Current Protocols in Molecular Biology, Ausubel, F. *et al.*, eds., John Wiley & Sons, *supra*) is used to identify the presence of a mutation associated with narcolepsy. For Northern analysis, a test sample of RNA is obtained from the individual by appropriate means. Specific hybridization of a nucleic acid probe, as described above, to RNA from the individual is indicative of a mutation in the HCRTR2 gene, and is therefore diagnostic for narcolepsy

For representative examples of use of nucleic acid probes, see, for example, U.S. Patents No. 5,288,611 and 4,851,330. Alternatively, a peptide nucleic acid (PNA) probe can be used instead of a nucleic acid probe in the hybridization methods described above. PNA is a DNA mimic having a peptide-like, inorganic backbone, such as N-(2-aminoethyl)glycine units, with an organic base (A, G, C, T or U) attached to the glycine nitrogen via a methylene carbonyl linker (see, for example, Nielsen, P.E. *et al., Bioconjugate Chemistry*, 1994, 5, American Chemical Society, p. 1 (1994). The PNA probe can be designed to specifically hybridize to a gene having a polymorphism associated with autoimmune disease. Hybridization of the PNA probe to the HCRTR2 gene is diagnostic for narcolepsy..

In another method of the invention, mutation analysis by restriction digestion can be used to detect mutant genes, or genes containing polymorphisms, if the mutation or polymorphism in the gene results in the creation or elimination of a restriction site. A test sample containing genomic DNA is obtained from the individual. Polymerase chain reaction (PCR) can be used to amplify the HCRTR2 gene (and, if necessary, the flanking sequences) in the test sample of genomic DNA from the test individual. RFLP analysis is conducted as described *(see* Current Protocols in Molecular Biology, supra). The digestion pattern of the relevant DNA fragment indicates the presence or absence of the mutation in the HCRTR2 gene, and therefore indicates the presence or absence of narcolepsy.

Sequence analysis can also be used to detect specific mutations in the HCRTR2 gene. A test sample of DNA is obtained from the test individual. PCR can be used to amplify the gene, and/or its flanking sequences. The sequence of the HCRTR2 gene, or a fragment of the gene is determined, using standard methods. The sequence of the gene (or gene fragment) is compared with the nucleic acid sequence of the gene, as described above. The presence of a mutation in the HCRTR2 gene indicates that the individual has narcolepsy.

Allele-specific oligonucleotides can also be used to detect the presence of a mutation in the HCRTR2 gene, through the use of dot-blot hybridization of amplified proteins with allele-specific oligonucleotide (ASO) probes (see, for example, Saiki, R. *et al.*, (1986), *Nature (London) 324*:163-166). An "allele-specific oligonucleotide" (also referred to herein as an "allele-specific oligonucleotide probe") is an oligonucleotide of approximately 10-50 base pairs, preferably approximately 15-30 base pairs, that specifically hybridizes to the HCRTR2 gene, and that contains a mutation associated with narcolepsy. An allele-specific oligonucleotide probe that is specific for particular mutation in the HCRTR2 gene can be prepared, using standard methods (see Current Protocols in Molecular Biology, supra). To identify mutations in the gene that are associated with narcolepsy, a test sample of DNA is obtained from the individual. PCR can be used to amplify all or a fragment of the HCRTR2 gene, and its flanking sequences. The DNA containing the amplified HCRTR2 gene (or fragment of the gene) is dot-

blotted, using standard methods (see Current Protocols in Molecular Biology, supra), and the blot is contacted with the oligonucleotide probe. The presence of specific hybridization of the probe to the amplified HCRTR2 gene is then detected. Specific hybridization of an allele-specific oligonucleotide probe to DNA from the individual is indicative of a mutation in the HCRTR2 gene, and is therefore indicative of narcolepsy.

Other methods of nucleic acid analysis can be used to detect mutations in the HCRTR2 gene, for the diagnosis of narcolepsy. Representative methods include direct manual sequencing; automated fluorescent sequencing; single-stranded conformation polymorphism assays (SSCA); clamped denaturing gel electrophoresis (CDGE) heteoduplex analysis; chemical mismatch cleavage (CMC); RNase protection assays; use of proteins which recognize nucleotide mismatches, such as *E. coli* mutS protein; allele-specific PCR, and other methods.

PHARMACEUTICAL COMPOSITIONS

The present invention also pertains to pharmaceutical compositions comprising nucleic acids described herein, particularly nucleic acids containing the HCRTR2 gene described herein. For instance, a nucleotide or nucleic acid construct (vector) comprising a nucleotide of the present invention can be formulated with a physiologically acceptable carrier or excipient to prepare a pharmaceutical composition. The carrier and composition can be sterile. The formulation should suit the mode of administration.

Suitable pharmaceutically acceptable carriers include but are not limited to water, salt solutions (e.g., NaCl), saline, buffered saline, alcohols, glycerol, ethanol, gum arabic, vegetable oils, benzyl alcohols, polyethylene glycols, gelatin, carbohydrates such as lactose, amylose or starch, dextrose, magnesium stearate, talc, silicic acid, viscous paraffin, perfume oil, fatty acid esters, hydroxymethylcellulose, polyvinyl pyrolidone, etc., as well as combinations thereof. The pharmaceutical preparations can, if desired, be mixed with auxiliary agents, e.g., lubricants, preservatives, stabilizers, wetting agents, emulsifiers, salts for influencing osmotic

pressure, buffers, coloring, flavoring and/or aromatic substances and the like which do not deleteriously react with the active compounds.

The composition, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents. The composition can be a liquid solution, suspension, emulsion, tablet, pill, capsule, sustained release formulation, or powder. The composition can be formulated as a suppository, with traditional binders and carriers such as triglycerides. Oral formulation can include standard carriers such as pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, polyvinyl pyrollidone, sodium saccharine, cellulose, magnesium carbonate, etc.

Methods of introduction of these compositions include, but are not limited to, intradermal, intramuscular, intraperitoneal, intraocular, intravenous, subcutaneous, oral and intranasal. Other suitable methods of introduction can also include gene therapy (as described below), rechargeable or biodegradable devices, particle acceleration devises ("gene guns") and slow release polymeric devices. The pharmaceutical compositions of this invention can also be administered as part of a combinatorial therapy with other agents.

The composition can be formulated in accordance with the routine procedures as a pharmaceutical composition adapted for administration to human beings. For example, compositions for intravenous administration typically are solutions in sterile isotonic aqueous buffer. Where necessary, the composition may also include a solubilizing agent and a local anesthetic to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the composition is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water, saline or dextrose/water. Where the composition is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

For topical application, nonsprayable forms, viscous to semi-solid or solid forms comprising a carrier compatible with topical application and having a dynamic viscosity preferably greater than water, can be employed. Suitable formulations include but are not limited to solutions, suspensions, emulsions, creams, ointments, powders, enemas, lotions, sols, liniments, salves, aerosols, etc., which are, if desired, sterilized or mixed with auxiliary agents, e.g., preservatives, stabilizers, wetting agents, buffers or salts for influencing osmotic pressure, etc. The agent may be incorporated into a cosmetic formulation. For topical application, also suitable are sprayable aerosol preparations wherein the active ingredient, preferably in combination with a solid or liquid inert carrier material, is packaged in a squeeze bottle or in admixture with a pressurized volatile, normally gaseous propellant, e.g., pressurized air.

Agents described herein can be formulated as neutral or salt forms. Pharmaceutically acceptable salts include those formed with free amino groups such as those derived from hydrochloric, phosphoric, acetic, oxalic, tartaric acids, etc., and those formed with free carboxyl groups such as those derived from sodium, potassium, ammonium, calcium, ferric hydroxides, isopropylamine, triethylamine, 2-ethylamino ethanol, histidine, procaine, etc.

The agents are administered in a therapeutically effective amount. The amount of agents which will be therapeutically effective in the treatment of narcolepsy can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the disease or disorder, and should be decided according to the judgment of a practitioner and each patient's circumstances. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture,

use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use of sale for human administration. The pack or kit can be labeled with information regarding mode of administration, sequence of drug administration (e.g., separately, sequentially or concurrently), or the like. The pack or kit may also include means for reminding the patient to take the therapy. The pack or kit can be a single unit dosage of the combination therapy or it can be a plurality of unit dosages. In particular, the agents can be separated, mixed together in any combination, present in a single vial or tablet. Agents assembled in a blister pack or other dispensing means is preferred. For the purpose of this invention, unit dosage is intended to mean a dosage that is dependent on the individual pharmacodynamics of each agent and administered in FDA approved dosages in standard time courses.

METHODS OF THERAPY

The present invention also pertains to methods of therapy for narcolepsy, utilizing the pharmaceutical compositions comprising nucleic acids, as described herein. The therapy is designed to replace/supplement activity of the hypocretin(orexin) receptor 2 in an individual, such as by administering a nucleic acid comprising the HCRTR2 gene or a derivative or active fragment thereof. In one embodiment of the invention, a nucleic acid of the invention is used in the treatment of narcolepsy. The term, "treatment" as used herein, refers not only to ameliorating symptoms associated with the disease, but also preventing or delaying the onset of the disease, and also lessening the severity or frequency of symptoms of the disease. In this embodiment, a nucleic acid of the invention (e.g., the HCRTR2 gene (SEQ ID NO:1)) can be used, either alone or in a pharmaceutical composition as described above. For example, the HCRTR2 gene, either by itself or included within a vector, can be introduced into cells (either *in vitro* or *in vivo*) such that the cells produce native HCRTR2 receptor. If necessary, cells that have been transformed with the gene or can be introduced (or re-introduced) into an individual affected with the disease. Thus, cells which, in nature, lack native HCRTR2 expression and activity, or have mutant HCRTR2 expression and activity, can be engineered to express

HCRTR2 receptors (or, for example, an active fragment of the HCRTR2 receptor). In a preferred embodiment, nucleic acid comprising the HCRTR2 gene, can be introduced into an expression vector, such as a viral vector, and the vector can be introduced into appropriate cells which lack native HCRTR2 expression in an animal. In such methods, a cell population can be engineered to inducibly or constitutively express active HCRTR2 receptor. Other gene transfer systems, including viral and nonviral transfer systems, can be used. Alternatively, nonviral gene transfer methods, such as calcium phosphate coprecipitation, mechanical techniques (e.g., microinjection); membrane fusion-mediated transfer via liposomes; or direct DNA uptake, can also be used.

The nucleic acids and/or vectors are administered in a therapeutically effective amount (i.e., an amount that is sufficient to treat the disease, such as by ameliorating symptoms associated with the disease, preventing or delaying the onset of the disease, and/or also lessening the severity or frequency of symptoms of the disease). The amount which will be therapeutically effective in the treatment of a particular disorder or condition will depend on the nature of the disorder or condition, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the disease or disorder, and should be decided according to the judgment of a practitioner and each patient's circumstances. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems.

The following Examples are offered for the purpose of illustrating the present invention and are not to be construed to limit the scope of this invention. The teachings of all references cited herein are hereby incorporated herein by reference.

EXAMPLES

EXAMPLE 1 Identification of the Human Narcolepsy Gene

A human BAC library (RPCI11 human male BAC library; see Osoegawa, K. *et al., Genomics 52*:1-8 (1998)) was used. Twenty primers, designed from the mRNA sequence of the HCRTR2 receptor, were employed to identify clones of interest. They are set forth in Table 1.

TABLE 1 Primers Used for Hybridization

#	Name	Primer Sequence	SEQ ID NO:
1	HCRTR2-1-F	TACTACTACTAGGCCACGCG	3
2	HCRTR2-1-R	ACACCAGGAGGAGAAAGCTAC	4
3	HCRTR2-2-F	ATCGCCTGTAAAGACAGCAAAG	5
4	HCRTR2-2-R	AAAGTTACTGAGCCAATGCCTC	6
5	HCRTR2-3-F	GAGAGGAGCTTGCAGCATTG	7
6	HCRTR2-3-R	AGGAATTCCTCGTCGTCATAGT	8
7	HCRTR2-4-F	GAAGAACCACCACATGAGGAC	9
8	HCRTR2-4-R	ATCACTTTGCAAAGGGACTGTC	10
9	HCRTR2-5-F	GTATGCAATCTGTCACCCTTTG	11
10	HCRTR2-5-R	AATGCAGGAGACAATCCAGATG	12
11	HCRTR2-6-F	CAGGCTTAGCCAATAAAACCAC	13
12	HCRTR2-6-R	GATAAGCCAACACCATGAGACA	14
13	HCRTR2-7-F	ACAGATCCCTGGAACATCATCT	15
14	HCRTR2-7-R	CTCGGATCTGCTTTATTTCAGC	16
15	HCRTR2-8-F	CCAATTAGCATCCTCAATGTGC	17
16	HCRTR2-8-R	GTGTGAAAAGGTAAACCAGGCA	18
17	HCRTR2-9-F	CTCAGTGGAAAATTTCGAGAGG	19
18	HCRTR2-9-R	GTTGCTGATTTGAGTGGTCAAG	20
19	HCRTR2-10-F	CTTTCTGAGCAAGTTGTGCTCA	21
20	HCRTR2-10-R	TACCAGTTTTGAAGTGGTCCTG	22

Initial Study with Large Membranes

Four out of 5 membranes having the whole BAC library, containing a total of approximately 160,000 BAC clones representing an approximately 10-fold coverage of the human genome, were used in hybridization studies with these primers. Hybridization was performed with a pool of all 20 primers described in Table 1.

5' End Labeling for Big Membranes

Oligonucleotides were labeled at the 5' end before hybridization, using fresh (less than one month old) [γ^{32}P]ATP (6000 Ci/mmole; 10 µCi/µl). The following protocol is adjusted for 4 membranes in 2 bottles, containing 2 membranes/30 ml of rapid hyb. Each. Briefly, a labeling mixture was made of DNA (8 pmol/µl) (10.0 µl of the primer pool), 10X buffer (12.0 µl), T4 PNK (10 u/µl) (6.0 µl), [γ^{32}P]ATP (30.0 µl, or 600 µCi), and water (62.0 µl) for a final volume of 120 µl. 20 µl of labeling mixture was used per 10 ml rapid hybridization reaction. Incubation of the labeling mixture was for 2 hours at 37°C, followed by transfer to ice, spinning down, and mixing with the rapid hybridization solution. The membranes were prehybridized at 42°C before the labeling mix was added. Sixty µl of the labeling mix was added to each of 2 big bottles containing 2 membranes and 30 ml of rapid hybridization solution.

Hybridization and Washing

The membranes were hybridized at 42°C overnight. After overnight, membranes were washed with 6x SSC, 0.1% SDS at room temperature; washed with 6x SSC, 0.1% SDS at 55°C in a shaking waterbath, repeated until the radioactivity of membranes was lower than 6k using 1x sensitivity; and washed with 6x SSC to remove the SDS. The washed membranes were put in a cassette for overnight exposure at -80°C with a MR single emulsion film. Positive clones were identified and gridded on small membranes.

Study of Positive Clones with Small Membranes

After growing the positively-identified clones on several small membranes (to get several copies of membranes containing the same clones), and washing the membranes, hybridization was performed using pairs of primers, instead of a total pool of primers as before. The total number of hybridizations was ten, using different primers against identical copies of membranes containing all positive clones from the first hybridization. The primer pairs are set forth in Table 2; primer numbers indicate the primers shown in Table 1.

TABLE 2 Primer Pairs Used for Hybridization

Reaction number	Primers Used
1	1 and 2
2	3 and 4
3	5 and 6
4	7 and 8
5	9 and 10
6	11 and 12
7	13 and 14
8	15 and 16
9	17 and 18
10	18 and 19

5' End Labeling for Small Membranes

Oligonucleotides were labeled at the 5' end before hybridization, using fresh [γ^{32}P]ATP (5000 Ci/mmole; 10 µCi/µl). Briefly, a labeling mixture was made of DNA (8 pmol/µl) (1.5 µl), 10X buffer (2.0 µl), T4 PNK (10 u/µl) (1.0 µl), [γ^{32}P]ATP (3.0 µl), and water (12.5 µl) for a final volume of 20 µl. Incubation of the labeling mixture was for 2.5 hours at 37°C, followed by transfer to ice, spinning down, and mixing with the rapid hybridization solution. Membranes were pre-wetted in 6X SSC, rolled in a pipette, and excess liquid drained prior to placing the membrane in the tube. Fifty ml Falcon (polypropylene) tubes were used as container for the hybridization. The membranes were prehybridized at 42°C before 20 µl of labeling mix was added to each tube.

Hybridization and Washing

The membranes were hybridized at 42°C overnight. After overnight, membranes were washed as described above. Four clones which were positive for primers designed using the 5' and 3' end of the mRNA were identified. Clone 403B19 was used to characterize the gene.

Sequencing of Narcolepsy Gene in Clone 403B19

Shotgun sequencing was used to obtain the gene sequence.

Preparation of DNA Samples

BAC DNA was isolated using the Plasmix kit from TALENT-VH Bio Limited. Thirty μg of isolated DNA was fragmented by nebulization: a nebulizer (IPI Medical Products, Inc., no. 4207) was modified by removing the plastic cylinder drip ring, cutting off the outer rim of the cylinder, inverting it and placing it back into the nebulizer; the large hole in the top cover (where the mouth piece was attached) was sealed with a plastic stopper; the small hole was connected to a 1/4 inch length of Tycon tubing (connected to a compressed air source). A DNA sample was prepared containing 30 μg DNA, 10 X TM buffer (200 μl), sterile glycerol (1 ml), and sterile dd water (q.s.) for a total volume of 2 ml. The DNA sample was nebulized in an ice-water bath for 2 minutes and 40 seconds (pressure bar reading 0.5). The sample was then briefly centrifuged at 2500 rpm to collect the DNA; the entire unit was placed in the rotor bucked of a table top centrifuge (Beckman GPR tabletop centrifuge) fitted with pieces of Styrofoam to cushion the nebulizer. The sample was then distributed into four 1.5 ml microcentrifuge tubes and ethanol precipitated. The Dried DNA pellet was resuspended in 35 μl of 1X TM buffer prior to proceeding with fragment end-repair.

Fragment End Repair, Size Selection and Phosphorylation

The DNA was resuspended in 27 μl of 1X TM buffer. The following materials were added: 10 X kinase buffer (5 μl), 10 mM rATP (5 μl), 0.25 mM dNTPs (7 μl), T4 polynucleotide kinase (1 μl (3 U/μl)), Klenow DNA polymerase (2 μl (5 U/μl)), T4 DNA polymerase (1 μl (3 U/μl)), for a total volume of 48 μl. The mixture was incubated at 37°C for 30 minutes, and then 5 μl of agarose gel loading dye was added. The mixture was then applied to separate wells of a 1% low melting temperature agarose gel and electrophoresed for 30-60 minutes at 100-120 mA. The DNA was then eluted from each sample lane, extracted from the agarose

using Ultrafree-DA columns (Millipore) and then cleaned with Microcon-100 columns (Amicon), precipitated in ethanol, and resuspended in 10 µl of 10:0.1 TE buffer.

Ligation

EcoRV-linearized, CIAP-dephosphorylated Bluescript vector was used as a cloning vector. The following reagents were combined in a microcentrifuge tube, and incubated overnight at 4°C: DNA fragments (100-1000 ng), cloning vector (2 µl (10 ng/µl)), 10X ligation buffer (1 µl), T4 DNA ligase (NEB 202L) (1 µl (400 U/µl)), sterile dd water (q.s.), for a total of 10 µl.

Transformation of Ligated Products

The ligation products were diluted 1:5 with dd water and used to transform electrocompetent TOP 10F cells (Invitrogen) using GenePulser II (Biorad; voltage, 2.5 W, resistance 100 ohm). Transformants were plated on LB plates with 50 µl of 4% X-GAL and 50 µl of 4% IPTG, and ampicillin. Transformants were grown overnight at 37°C, white colonies were picked, grown in a culture of 3 ml LB liquid media plus 200 µg/µl ampicillin for 16-20 hours with shaking. DNA was isolated from the liquid cultures using Autogen 740 Automatic Plasmid Isolation System.

Cycle Sequencing of Isolated Plasmid DNA

Isolated plasmids were then sequenced using the M13 primers: M13-forward (SEQ ID NO:23) TGTAAAACGACGGCCAG; and M13-reverse (SEQ ID NO:24) CAGGAAACAGCTATGAC. For the sequencing reaction, 2.5 µl plasmid template was mixed with 4 µl Big Dye Ready reaction mix (ABI), 1 µl of 8 pM M13 primer, and 2.5 µl dd water. For cycle sequencing, 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds, and 60 °C for 4 minutes were performed, followed by holding at 4°C. The cycle sequencing reaction products were cleaned by spinning through Sephadex G-50 columns. The eluted cycle sequencing products were then dissolved in 3 µl formamide/dye and 1.5 µl of sample was loaded on ABI 377 automated sequencers. The data was analyzed using Phred and Phrap (Ewing, B. *et al., Genome Res. 8*:175-

185 (1998); Ewing, B. and Green, P., *Genome Res. 8*:186-194 (1998)), and viewed in Consed viewer (Gordon, D. *et al., Genome Res. 8(3)*:195-202 (1998)).

Analysis of Gene Structure

The *hcrtr-2* gene maps to chromosome 6p11-q11. A total of 168,575 base pairs of contiguous sequence was generated for 403B19 which contained all of the *hcrtr-2* gene. Comparison of the cDNA sequence of *hcrtr-2* (Accession number GI:6006037) and the genomic sequences generated allowed deduction of the intron/exon organization of the gene. The gene contains 7 exons which cover 108,439 bp. The first 10 Gs in the mRNA sequence for *hcrtr-2* were not found in the genomic sequence. It is likely that these Gs were an artifact.

The splice junctions of the *hcrtr-2* gene are set forth in Table 3, and the intron sizes are set forth in Table 4. Exon sequences are represented in uppercase and introns in lowercase. All splice sites conform to the consensus GT-AG rule. SEQ ID NOs are given in the column immediately following each site.

Table 3 Splice Junctions of *hcrtr-2*

	Splice Donor Site	SEQ ID	Splice Acceptor Site	SEQ ID
Hcrtr-2 exon1-2	TCCTGGgtgagt	25	aattagTTTGTG	26
Hcrtr-2 exon2-3	CTACAGgtaatt	27	ctctagACCGTG	28
Hcrtr-2 exon3-4	GGGGTGgtaagt	29	tcctagGTGAAA	30
Hcrtr-2 exon4-5	CGACAGgtatat	31	tttcagATCCCT	32
Hcrtr-2 exon5-6	AAAGAGgtaaaa	33	ctgcagAGTATT	34
Hcrtr-2 exon6-7	TCAGTGgtgagt	35	tgccagGAAAAT	36

Table 4 Intron Sizes of *hcrtr-2*

Intron	Nucleotides
Intron 1	73,848
Intron 2	6,322
Intron 3	8,327
Intron 4	13,618
Intron 5	2,730
Intron 6	1,779

The exons do not clearly respect the domain structure of this seven membrane domain G protein linked receptor. Five of the transmembrane regions are by themselves within one exon, two of the transmebrane segments are broken up by introns, and two transmembrane segments fall within the same exon. A survey done one year ago on mammalian G-protein coupled receptors (GPCRs) sequences in GenBank revealed that over 90% of GPCRs genes were intronless in their open reading frame (ORF) (Gentles, A.J. and Karlin, S., *Trends Genet. 15*:47-49 (1999)). Comparison of the intron/exon boundaries of *hcrtr-2* and the genes coding for their most related GPCRs based on sequence similarity showed that the location of the intron/exons boundaries with respect to the transmembrane domains is only partially conserved among the receptors (Sakurai, T. *et al.*, *Cell 92*:573-585 (1998)).

Computer analysis of sequence data

Analysis of the genomic sequence of *hcrtr-2* using the program RepeatMasker (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker) showed that the sequence containing the *hcrtr-2* genomic sequence is 38.27% repeat sequences and the GC content is 35.3%.

The sequences of the genes were analyzed using the program GeneMiner (Óskarsson and Pálsson, unpublished), which combines the results of 5 exon prediction programs; FGENE (Solovyev, V. and Salamov, A., *Ismb 5*:294-302 (1997)), Genscan (Burge, C. and Karlin, S., *J. Mol. Biol. 268*:78-94 (1997)),

HMMgene (Krogh, A., *Ismb 5*:179-186 (1997)), MZEF (Zhang, M.Q., *Proc. Natl. Acad. Sci. USA 94*:565-8 (1997)) and Xpound (Thomas, A. and Skolnick, M.H., *IMA J. Math Appl. Med. Biol. 11*:149-160 (1994)). For *hcrtr-2,* 3 out of 5 programs predicted the 3' end of exon 1, only one program predicted the 7th exon and for the internal exons, there were at least two programs that predicted each of them exactly or in part.

The promoter sequences of the genes have not yet been characterized. The Promoter Prediction by Neural Network (http://www.fruitfly.org/seq_tools/promoter.html) predicted promoters that are at least 140 bp upstream of the 5′UTR of *hcrtr-2*, indicating that either a part of the 5′UTR is missing in the published mRNA sequence or the real promoters are not detected by the program.

Analysis of Population for Polymorphisms

Each exon and its flanking intronic sequences of the *hcrtr-2* gene was analyzed in nucleic acid samples from 47 patients and 75 control individuals. The patient population consisted of patients of Icelandic and US origin. The control population consisted of Icelandic controls, CEPH (Centre d'Etude du Polymorphisme Humain) individuals from Utah and France, and US samples of various ethnic origins. The African-American/Caucasian ratios were similar between patients and controls. All narcoleptic subjects complained of excessive daytime sleepiness (EDS). Approximately 66% of the patients had cataplexy, 24% did not and 10% did not have attainable records of cataplexy status. Narcoleptic subjects without cataplexy had Multiple Sleep Latency Tests showing mean sleep latencies of less than 10 minutes and REM sleep in at least 2 naps. Subjects did not take any drugs affecting sleep for at least 10 days before their sleep studies.

To analyze the nucleic acids, DNA from patient and control blood samples were isolated by the method of Kunkel (Kunkel, L.M. *et al., Proc. Natl. Acad. Sci. USA 74*:1245-9 (1977)). Briefly, white blood cells were lysed in a sucrose lysis buffer, and proteinase K treated; the DNA was then extracted using phenol-chloroform/isoamylalcohol and then ethanol precipitated. Patient samples that were

received in the form of nuclei pelleted through sucrose buffer were resusupended in lysis buffer (100 mM NaCl2; 10 mM TrisHCl, pH 8; 25 mM EDTA pH 8; 0.5% sodium dodecyl sulfate; 0.1 mg/ml proteinase K) at 55°C for 4-6 hours followed by classical phenol-chloroform extraction and ethanol precipitation (Sambrook, J. *et al.*, *Molecular Cloning, A Laboratory Manual* (1989)). Samples were incubated at 55°C after isolation for the inactivation of DNAse to prevent the degradation of DNA. Concentration of the isolated DNA was determined by spectrophotometric analysis at 260 nm (Sambrook *et al.*,using GeneQuant (PharmaciaBiotech), and samples diluted with sterile distilled water to a 20 ng/µl working solution.

Primers were designed from intronic sequences flanking the exons of the hypocretin receptor-2 (*hcrtr-2*), using either primer design programs available at primer3 at the Whitehead Institute (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi) or primers for the worldwide web (http://williamstone.com/primers/javascript/). The primers are shown in Table 5.

Table 5 Primers Used to Amplify Nucleic Acid Fragments for Analysis of *hcrtr-2* Gene

EX-ON	#	Primer Sequence	Sense/ Antisense	External/ Nested	SEQ ID.
1	1	TTTCTTCAGCTTCAGCTCTCCCTCAGC	S	E	37
1	2	TTCAGCTCCGAAGCAGATGACCAGTTG	A	E	38
1	3	TTCAGCTTCAGCTCTCCCTCAGCGAGG	S	N	39
1	4	CGAAGCAGATGACCAGTTGCGACAAGG	A	N	40
1	5	CTTTCCCACCGCAAATCACCAGTGCTC	S	E	41
1	6	ATTTTATTAGAAAACCCCATCCGAGAG	A	E	42
1	7	TTCCCACCGCAAATCACCAGTGCTC	S	N	43
1	8	TATTAGAAAACCCCATCCGAGAGCAG	A	N	44
2	9	GCATGTACTTAGCATTCACACAGATTG	S	E	45
2	10	TCTAATGATGATTTGGCAGTTCATTGC	A	E	46
2	11	TAGCATTCACACAGATTGACAGATTCA	S	N	47
2	12	CAGTTTGTCAATGCCTTAGGCAAATAT	A	N	48
3	13	TTTGGCAGCTTTGAATTTGCTTATATG	S	E	49
3	14	GCTCTTGCAAAACTGTATTCACAAATG	A	E	50
3	15	CAGCTTTGAATTTGCTTATATGTTGTG	S	N	51
3	16	TTGCAAAACTGTATTCACAAATGTCAA	A	N	52
4	17	TCCCCTTTGCATACATAATATGACAATG	S	E	53
4	18	AAAAAGCACAGACAAAATATTTGGAAGG	A	E	54
4	19	ATGCACTTTGAAGAAAAGCATTGACATG	S	N	55
4	20	AAGCACAGACAAAATATTTGGAAGGAAT	A	N	56
5	21	CTCAGGCGTCTGGAAGCCTTTCCTTAC	S	E	57
5	22	TTAAAGGCTGTTCGCCTTACCTGCTGG	A	E	58
5	23	GGCGTCTGGAAGCCTTTCCTTACTGTG	S	N	59
5	24	CTGAGTCATCTGGCCTGACAAGGTATC	A	N	60
6	25	GGGTCAGAAACCAATCTGTGGTCAATTC	S	E	61
6	26	AGTTGAAGAGTGTTCATTGATTCCTCATCC	A	E	62
6	27	AGAAACCAATCTGTGGTCAATTCCTGCAAC	S	N	63

EX-ON	#	Primer Sequence	Sense/ Antisense	External/ Nested	SEQ ID.
6	28	TGAAGAGTGTTCATTGATTCCTCATCCTTG	A	N	64
7	29	GAGTCTACCAAGCTTCCAATAAACTCA	S	E	65
7	30	GGATAGTTTTACTCAGGTATCCTTGTCA	A	E	66
7	31	CAAATCAGCAACTTTGATAACATAT	S	N	67
7	32	GTATCCTTGTCATATGAATAAATATTCTAC	A	N	68
7	33	CACTCAAATCAGCAACTTTGATAAC	S	E	69
7	34	GTGAGAGATTAAAATAACAAGGGAT	A	E	70
7	35	CAAATCAGCAACTTTGATAACATAT	S	N	71
7	36	TGTTTAAACATTTAATTGACACACA	A	N	72
7	37	TTCATATGACAAGGATACCTGAGTAAA	S	E	73
7	38	GTGAAATAGCCTGAAATAAGCTCAA	A	E	74

PCR reactions were done in 20 µl reactions using 40 ng genomic DNA, 0.2 mM solution of the four dNTPs, 0.35 µM of each primer (TAGCopenhagen), 2.5 mM MgCl2 (Perkin Elmer), 1x PCR Buffer (Perkin Elmer) and 0.5 U Amplitaq gold (Perkin Elmer). The primers were used to amplify the fragments by PCR cycling at 95°C for 12 min and subsequently 30 cycles of 95°C for 30 sec, 55-62°C for 30 sec and 72°C for 1 min. The PCR products were prepared for cycle sequencing by incubation with Shrimp alkaline phosphatase (Amersham) and exonuclease I (Amersham) at 37°C for 15 min. After the inactivation of the enzymes the products were subject to cycle sequencing using BigDye Ready Reaction mix (Perkin Elmer) and subsequently run on ABI Prism 377 Automated DNA sequencers. The raw data were basecalled and sequences assembled using the Phred and Phrap software, respectively (Ewing, B. *et al., Genome Res. 8*:175-185 (1998); Ewing, B. and Green, P., *Genome Res. 8*:186-194 (1998)). The Consed viewer was used to analyze the sequences (Gordon, D. *et al., Genome Res. 8(3)*:195-202 (1998)). Expansion of a T-stretch in the 3' untranslated region (UTR) of exon 7 of *hcrtr-2* was investigated by amplifying a fragment containing the stretch with a fluorescently labelled primer

pair using the conditions described above. The PCR product was dissolved in formamide/dye solution and run on ABI Prism 377 Automated DNA sequencers as described above. Allele calling was done using TrueAllele and editing was done using DeCODE-GT (Palsson, B. *et al., Genome Res. 9*:1002-1012 (1999)).

A total of nine single nucleotide polymorphisms were identified, 7 in exons and 2 in an intronic sequence near an exon. The polympoprphisms are shown in Table 6. The base number is according to the mRNA sequence (Accession number GI:6006037). For those polymorphisms marked with an asterisk (*), the polymorphism is located 5' of the corresponding exons; the numbers indicate the distance into the introns.

Table 6 Single Nucleotide Polymorphisms in *hcrtr-2*

Location	cDNA base #	Nucleic Acid Change
Exon 1	352	C-T
Exon 1	355	C-A
Intron1	-26*	C-A
Exon 5	1,170	G-A
Exon 5	1,177	C-A
Exon 5	1,201	G-A
Exon 5	1,246	G-A
Exon 5	1,266	G-A
Intron 6	-87*	G-A

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the spirit and scope of the invention as defined by the appended claims.

CLAIMS

What is claimed is:

1. Isolated nucleic acid molecule comprising the nucleic acid having SEQ ID NO:1.

2. A DNA construct comprising the isolated nucleic acid molecule of Claim 1 operatively linked to a regulatory sequence.

3. A recombinant host cell comprising the isolated nucleic acid molecule of Claim 1 operatively linked to a regulatory sequence.

4. A pharmaceutical composition comprising a nucleic acid comprising the isolated nucleic acid molecule of Claim 1.

5. Isolated nucleic acid molecule comprising the nucleic acid having SEQ ID NO:1 with one or more of the nucleic acid changes shown in Table 6.

6. A method of diagnosing narcolepsy in an individual, comprising detecting a mutation in the gene encoding hypocretin (orexin) receptor 2, wherein the presence of the mutation in the gene is indicative of narcolepsy.

7. A method of treating narcolepsy in an individual, comprising administering to the individual an isolated nucleic acid of Claim 1 in a therapeutically effective amount.

```
LOCUS           __________      168,575 bp    DNA      PRI     20-OCT-1999
DEFINITION      Human hypocretin (orexin) receptor 2 (HCRTR2) gene, complete cds.
ACCESSION       __________
NID             __________
VERSION         __________
KEYWORDS        .
SOURCE          human.
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1-168,575)
AUTHORS         __________
  TITLE         Direct Submission
 JOURNAL              Submitted (__________) deCode Genetics, Inc., Lynghals 1,
                      IS-110   Reykjavik, Iceland.
FEATURES              Location/Qualifiers
     source           1.. 168,575
                      /organism="Homo sapiens"
                      /db_xref="taxon : 9606"
                      /chromosome="6"
                      /map="6p11-q11"
                      /clone="BAC 403B19"
     gene             1..129,305
                      /partial
                      /gene="HCRTR2"
                      /note="OX2R"
                      /db_xref="LocusID:3062"
                      /db_xref="MIM:602393"
     exon             20,867..21,403
                      /gene="HCRTR2"
                      /number=2
     CDS              join(21,181..21,403, 95,252..95,430, 101,753..101,996, 110,324..110,439,
                      124,058..124,278, 127,009..127,130, 128,910..129,139)
                      /gene="HCRTR2"
                      /note="HCRTR2 exons defined by comparison to mRNA sequence (NM_001526)"
                            /product="HCRTR2/orexin 2 receptor"
                      /db_xref="LocusID:3062"
                      /db_xref="MIM:602393"
                      /protein_id="NP_001517.1"
                      /db_xref="PID:g4557639"
                      /db_xref="GI:4557639"
                      /translation="MSGTKLEDSPPCRNWSSASELNETQEPFLNPTDYDDEEFLRYLW
                      REYLHPKEYEWVLIAGYIIVFVVALIGNVLVCVAVWKNHHMRTVTNYFIVNLSLADVL
                      VTITCLPATLVVDITETWFFGQSLCKVIPYLQTVSVSVSVLTLSCIALDRWYAICHPL
                      MFKSTAKRARNSIVIIWIVSCIIMIPQAIVMECSTVFPGLANKTTLFTVCDERWGGEI
                      YPKMYHICFFLVTYMAPLCLMVLAYLQIFRKLWCRQIPGTSSVVQRKWKPLQPVSQPR
                      GPGQPTKSRMSAVAAEIKQIRARRKTARMLMVVLLVFAICYLPISILNVLKRVFGMFA
                      HTEDRETVYAWFTFSHWLVYANSAANPIIYNFLSGKFREEFKAAFSCCCLGVHHRQED
                      RLTRGRTSTESRKSLTTQISNFDNISKLSEQVVLTSISTLPAANGAGPLQNW"
     exon             95,252..95,430
                      /gene="HCRTR2"
                      /number=3
     exon             101,753..101,996
                      /gene="HCRTR2"
                      /number=4
```

FIG. 1A

```
exon            110,324..110,439
                /gene="HCRTR2"
                /number=5
exon            124,058..124,278
                /gene="HCRTR2"
                /number=6
exon            127,009..127,130
                /gene="HCRTR2"
                /number=7
exon            128,910..129,305
                /gene="HCRTR2"
                /number=8

BASE COUNT    55,308 a  29,672 c  29,838 g  53,757 t

CGACTTGATTTTATTTTTTGCATATGGATATCCAGTTTTCACAGCACTGCTTGTTACCCT
CAGCAAAGAACAGTTGTCTGTAAATTCATGGGTTTATGTCTAGGCTCTCTGTTCTGTTCT
ATTGGTCAACATATGGTCATATATCACTTAACTGCAGGGAAGGGATACATTCTGAGAAAT
GCATTATTACATGATTTCATCATTGTGCAAACACTATAGAGTGTAGTTACAGAAACCTAG
TATCTCTAGCTGTGTTCTTATGATTCAAATTTGCTTTGGTCATTTGAGATCCATACTGGT
GGAGTCTAATTATTCAAAACTAGGGAAAACAGACAAACAGAAAAAACTAAGACCAAGTTA
GCAGAAGAAAGACAATAACAAAGGTTAGATCAAAAATAAATAATATAGAGAATGAAAAAA
TTAGAAAAAGTGGACAAAACTACAATGTACTTTTTTGAAAAGACAAACAAAATTAACAAA
CCCTTACCTTGACTAAAAAAAGAGACTCAAATAAATAAAATTGGAAATGAGACAGGAGAC
ATTACAATTGATGTTAACAAAAAGATCATAAGGTACTATTATGAACAACTATACACCAAT
AAATTGGACAACCTAGAAAAAAAATGGATAAATTCCTAGAAATACACAGTCTATCAAACT
GAAACAAGAAGAAATAGAAAGCCTGAACATACCAGTAACAACCAAGGAGACTGAGTAAAT
AATCAAAAACCTCCCAAGAAGAAGAGTCTAGGACCAGAAGTCTTCACAAATGAATTCTAC
CAAACATTTAAAGTATTAATGCCAATCATTCATTCTTATACTCTTCCAAAAAGAAAGAGG
GAATATTTTCAAACTCATTTTATGAGGCCAGCATTATTCTGATACCAAAACTACGCAAAA
ATACTACAAGAACATAAAAACTACAAATGTGGGAATTATCATGTATACATATGCAAAAAT
CCTCAGTAAAATCCTAGCAAACTAAATTCAACAGTACATTAAAAAGATCATATAGCATGA
CCAGTGAAATTTCTCCTTAGGACGCAAGGATAAGTCAACATATAAAATTGAATGTGATAT
ACCACTTTAACAAAATGAAGGATAAAAATCATATGATCATCTGAATAGATGCAGAAAAAG
CATATAACAAACTTTGACGTTGTTGAGAAATTGAAAGCTTTTCTCTAAGATCAAGAACAA
AGCAAGGATGCCCATTCTTGCTTCTATTCAGCATAGTGCTTGAAGTCCTAGTCTGGACAA
TTGGGCAAAAATAAATAAATAAATAAATAGATAAATAAATAAATAAATATAAATAAATAA
ATAAAATCCACCAAATTGGAAAGGAAGAAGTGAAATTACCTCTGTTTGTAGATGAGCTGA
TCTCATGTGTAGACAACCTTAAAGATTCCACAAAAACAAACAAACACACAAACAAACAAA
ATAGCTAGAGCAAAGAAATGAATTCAGTACAGTTGCAGAATGCAAAATCAGTATACAAAA
AGTACTTGTAATTCTATATAATAGCAACAAACTATTTCATAAGGAAATTAAGGAAACAAT
CCCCATTACAATAGCATCATAAATAATAAATCTTAAGAACAAATTTAACCAAGGAGGTGA
AAGACTTGTGTACTGAAAACTATAAAATGCTGATAAAAAAATTAAAGAAGATACAATAAA
TGGAAGATATTCCATGCCATGGTTTGGAAGAATTAATATTGCTAAATGTACATACTACCC
AAAACAACCTGTAGAGTCAATGCAATCCCTATCAAAATACCAATATTTTTTTTTTTACAG
AAATGAAAACACAATCTTAACACTATTTAAACCAATTAAACAAACCTATGATTTCAATTT
GGTCAAATGTGTTAGAATGGATTTCCTTTTATTGTTTTGAACTTGTCTCTTCCAAATTTC
AAAGCCTGGTTCCTAATTTTTACTTGAAATACCAAATAACAAACCCACTTAATGAGCTCT
GAGCCAGTTTTAGTAGCCAAACTTGTATTTAAATAGTGTGTTACATATTTGCACAAAAAG
CCAACGGAGTCTAAATCAACACTAATTCACATCATTACTAGCAATCTAAAACATCAGATG
ATAATTTTGCTGTTGTCTTTCAGGCAAGATATTCAACCATTGGTATTAAATGTTTTATAT
GAATGTGCGGTGTTTTATTTCAGAAACACTTCTCTGAATTCCCAAGGCCTAAGAGCTATT
CATCATAGAGGTTTGTGGAGGCGGTAGTTAGACATTTTCTACATGCATAATGTTAATTCA
TTCAAACATTATAGAAAAAAAGTTTGTAAAGAAGTTAATTTTCAAGGTGACAAAAAAATC
AGATTGAATCATGTTTATTTTATTTCAATTTAAACTCGTTGGCTATCTTAGGAAATTCAC
ATTGTTTTTGAAGAATATATGAACAAAGTTTGATTCATCTTATCTATATAAGCATGAGAG
```

FIG.1B

AAATACCTAGATGAGAGTGAAAAATGACTAATTTTGTGACCATTGTTATATCATAGATTA
ACTTGTTCTCTTCTACTTCTAAGCTGTGTGATCTTGAAAAGTCATCTAAACTTCAGGTAC
CATCCTCACTTGCAAAATGAGGGGAAAAACCCCAGCACCTTTAATATGGTGTTATGTGGA
TGAAATAAGTTAATATATATTAAGTGCTTAGGTTTCATGCACTTTCTATATAGTATTAAT
AATATTATTGTTACATTATTATAGTTACATTATTATTTTTATTAATATTATTGGAACATG
AATGGAATTGTTGTGGCTCATTTTAAAGATGCTGCAATGGAGACCAAGAGAAATTAAGTA
TAATAATCCTAGTTAAGGTACAGCCATTTCTAATTACATTTTCCAACTGCTGCTTTTACT
CCTAGCACTCACACCAATTCTTCTCATAATCTAATAAATACTGAAATTTAAAACTTATAA
AGAACACATAATAATCTTATTTAATTATCACAACAATTTCTGTGGAGTTACTATTAATCC
AGAGATGAAGAATCTAAAACTCTAAATTATCAAGCAACTATTCCAGCTTTAAACAACAGT
AAAACTGGAATTAAAACTAGAGTTTCTTTATGAGGCCAGTATTACTCTTATACTAAAGCA
AGACAACTGTCTCTCTCTCACTCTCTCTTTCTCTCTCTCTGAGACACACACACACACACA
CAAGCACACACACACACACACACCAGATCAATATAACTTATGGATGTTGATGCAAAAATTTT
CAAAAAATTAATAGCAAATCGAATCCAGCAGTATTTTAAAGGACTATACACCATGGACAA
ATGGGGTTTATTCCTGGGATATAAAGTTGGCTAAACTTAATGAAAATCAACCAGTGCAAT
AAATAATAGTAATTAAAAAAACATAATTATCTCACTAGATGTACAAAAAATGACAAGATC
CAACATGTTTTCAATATAAAAGCATTCCACAGACTAGGAATAGAAGGGAACTTCCTCAAC
TTTACAAAGAACATCTACAACGAAACCACAGCTAACATCATATTTAATGGTGGAAGACTG
AAATCTCTGAATATTTTCCCCTAAGATCAGAAAAAGACAAAGAGGTCTACTAATTCTATT
CAACATTGGAAAGATAGTTCTAGTCAATTAATCATTAAAAAAAGGCATTTAGATTAGAAA
GGAAGTAAAATTACCTCTGGCAGATGACATAATCTTATACATGGAGAACTCCTAGAGATT
ACATACACACTCACAACTACCAGAGTTAATAAATGGGTTCTACAAAGTTGCAGGATACAA
TATCAATTCTCAAAAAACACTTGTATTTCTATACACTAGCAAATAACTCTGAAAATGAAA
ATAACAAAACAATTCCACTTGCAAGAGCATCAAAAAAGCATGAAAATCTTAGGAAGGAAT
TTACCAAGAATGTACAAGTTTTATGTACTGAAAAATAAAAAATGTCATTAAAAATAGTTA
AAGAGAATCTAAGCACATTGCAGTTTTCTGACTCCAGGCCCGGGCTCTTGGATGGCATCT
CTGGATCCACTCAGGACCAGGGAGAACTTGTTGCCCTGAAGGGAAGGACACAAGTCTGAC
TGGCTTTACCACCTGCTGATTGTAGAAGCCTAGGGCCTTCAGGGAACACAGGTGGTAGCC
AGATAGCAGTTACCATGGGCATTAGGCATGACCCAGTGCTATGTTGGCTTCTAGTTTGAC
CCAGCACAGCCCAAGGGTGGTAACCACATGGGTGCTTGTGTCACCCCTCCTTAAGTTCCA
GGCAGCCAGCAAAGAGAGAGTGACTCTGTTTGGGAGAAAGTAAGGGAAGAGAATAAAAGT
CTCTGTTGGTAATACAAGGAATTCTTCCAGATCTTATCCAAGACCTCTATGAATCTGCAA
CAGCCAAAGCATTATTAGTTTTCAGGTTTCCCCAGTGCAGATATGACTGCAATGATCAAA
AACTTAGATTATAACACTCAAGTCCCATTTGATACCTGAAAAGCTTTCCAAGAAAGATAG
GCACAAACAAGTTTGGACTGGGAGGACTACAATAAATACCTAACTTCTCAATGCCCAGAA
ACTGATGAACATCCACAAGCTTTAAGACCATCCAGGAGGTGGCTGGCAAGATGGCTGAAT
AGGAACAGCTCTGGTCTGCAGCTCCCAGTGAGATCAGTGCAGAAGGTGGTGGTTTCTGCA
TTTCCAACTGAGGTACCCAGCTCCTCTCATTGGGACTGGTTAGATAGTGAGTGCAGCCCA
CAGAGGGTGAGCCAAAGCTGGATGGGGTGTCACCTCACTGGGGAAGCACAAGGGGATGGG
GAACTCCCTCCCCTAGCCAACGGAATCTGTGAGGGACTGCCATGAGGGATGGTGCATTCT
GGTCCAGATACTATGCTTTTCCCATGTTCTTCACAACCCTCAGGCCAGGAGATTCCCTCG
GGTGCCTACACCACCAGGGCCTTGGGTTTCAAGTACAAAACTGGGTGGATCTTTGGGCAG
GCACCGAGCTAGCTGCAGGAGTTATTTTTCATACCCCAGTGGTGCCTGGAATGCCAGTGA
GACAGAACCATTCACTCTCCTGGAAAGGGAGCTGAAGCCAGGGAACCCAGTGGTCTAGCT
CGGTGGATCCCACTCCCATGGAGGCCAGTAAGCTAAGATCCACTGGCTTGAAATTCTCAC
TGCCAGTGCAGCAGTCTGAAGTCAACCTGGGATGCTTGAGCTTGGTGGAGAGAGGGACGT
CCACCATTACTGAGGTTTGAGTAAGCAGTTTTCCCCTCACAGTGTAAACAAAGCCACTGG
GAAGTTAAAGTAGGTGGAGCCCACGACAGTTCGGCAAAGCCACTATAGCCAGAATGCCTC
TCTAGATTCCTCCTGTCTGGGCAGGGCTTCTCTGAAAGAAAGGCAGCAGCTGCGGTCAGG
AGCTTATAGATCAAACTCCCATCTCCCTGGGACAGGGCACCTGGGGAAAGGGGCAGCTGT
GGGTGCAGCTTCAGCAGACTTAAATATTGCCGCAAGCTGACTCTGAAGACAGCAGGGGAT
CTCCCAGCACAGCGCTCGAGCTCTGCTAAGGGGCAGACTGCCTCCTCAAGTGGGTCTCTG
ACCCCTGTGTCTCCAGACTGGGAGACACCGCACAGCAAGGGTCGACAGACACCTCATACA
GGAGAGCTCCGGCTGGTATCTGGTGGGTGCCCCTCTGGGACAAAGCTTCGAGAGGAAGGA
GCAGGCAGCAATCTTTGCAGTACTGTAGCCTCTACTGGTGATACCCAGGCAAATAGGGTC
TGACGTTGACCTCCAGCAAACTCCAGCAGACCTTCAGCAGACGGGCCTGAGTGTAAGAAG
GAAAATTAACAAACAGAAAGGAATAGCATCAACATCAAAAAAACAAAAACAAAAACAAAA

FIG. 1C

ACAAAAACAAAAACAGCACATCCGCACAAAAACCCCATCTGAAGGTCACCAACACCAAAT
ACCAAAGGTAGATAAATCCACAAAGATGGGGAAAAACCAGCACAAAAAAGCTGAAAATTC
CAAAAAACAGAATACCTCTTCTCCTCCAAAGGATCACAATTCCTCACCAGCAAGGGGACA
AAACTGGACAGAGAATGAGTTTGATGAATTGACAGAAGTAGGCTTGAAAAGGTGGGTAAT
AAACTCCTCTGAGCTAAAGGAGCATGTTCTAACCCAATGCAAGGAAGCTAAGAACCTTGA
AAAATGGTTAGAGTAATTGCTAACTAGAATAACCAGTTTAGAGAAGAGCATAAATGACCT
GATGGAGCTGAAAACTATAGCACAAGAACTTCGTGCAGCATACACAGGTATCAATATCCA
AATCGATCAAGCAAAGAAAAGAATATCAGAGATTGAAGATCAACTTAATGAAATAAAGTG
TGAAGACCAGATTAGAGAAAAAAGAATAAAAAGGAATGAACAAAGTCTCCAAGAAATATG
GGAATATGTGAAAAGACTAAACCTACATTTGATTAGTGTACCTGAAAGTGACGGGGAGAA
AGGAATCAAGTTGGAAAACATTCTTCAGGATATTATCCAGGAGAACATCCACAACCTAGC
AAGACAGGCCAACATTTAAATTCAGGAAATACAGAGTACATCACAAAGATACTCCTCGAG
AAAAACAACCCCAAGACACATAATTGTCAGATGCACCAAGGTTGAAATACAGGAAAAAAG
TTAAGGGCAGCCAGAGAGAAAGGTCGGGTTACCCACAAAGGGAAGCCCATCAGACTAACA
GTGGATCTCCCTGCAGAAACCCTACAAGCCAGAAGAGAGTGAAGGCCAATATTCAACATG
CTTTAAGAAAAGAATTTTCAACCCACAATTTCATATCCAGCCAAACTATGCTTCATAGTG
AAGGAGAAATAAAATCCTTTACAGACAAGCAAATGCTGAGAAATTTTGTCACCACCAGGC
CTGCCTTACAAGAGCTCCCGAAGGAAGCACTAAATATGAAAAGGAAAAACCAGTATCAGC
CACTGCAAAAACATATGAAATTGTAAAGACCATCAACACTATGAAGAAACTGCATCAACT
AATGGGCAAAATAACCAGCTAGCATTATAATGACAGGATCAAATTCACACATAACGATAT
TAACCTTAAATGTAATAGGCTAACTGCCCCAATTAAGAGACACAGACTGGCAAATTGGAT
AGAGAGTCAAGACCCAACAGTGTGCTGTATTCAGGAGTCCAATTCATGTGCAAAGATACA
TATAGGCTCGAAATAAAGGGATGGAGGAATATTTACTAAGCAAATGGAAAGCAAAATAAA
GCGGAGGTTGCAATCCTAGTCTCTGATAAAATAGACTTCAAACCAACAAAGATCAAAAGA
GACAACAAAGGGCATTACATAATGATAAAGGGATCAATGCAACAAGAACAGCTAGCTATC
CTAAATATATATGCACCCAATTCAGGAGCACACAAATTCATCAAGCAAGTTCTTAGAGAC
CTATAGAGACTTAGACTCCCACGTAATAATAGTGGGAGACTTTAACACCCCACTGTCAAT
ATTAAACAGATCAATGAGACAGAAAATTAACAAGTACATTCAGGACTTGAACTCAGCTCT
GGACCAAGCAGGCCTAATAGACATCTATAGGACTCTCCACCCCAAATAAATAGAATATAC
ATTATTCTCAGCACCACATTGCACTTATTCTAAAATTGACCACATCATTGGAAGTAAAAG
ACTCCTCAGCAAATGCCAAAGAACTAAAATCATAACAAACAGTCTCTCAGACCACAGTGC
AATCAAATAAGAGCTCTGGAATAAGAAACTCACTCAAAACCGCACAACTACATGGAAACT
GAACAACCTGCTGCTGAATGACTACTGGGTAAATAATGAAATTAAGGCAGAAATAAATAA
GTTACTTGAAACCAATGAGAACAAAGACACAACATACCAGAATCTCTGGGACACAGCTAA
AGTAGTGTTTGGAGGGAAATTCATAGCACTAAATGCCCACACGAGAAAGTGGGAAAGATC
TAAAATCAACACCCTAACATCACAATGAAAAGAACTAGAGAAGCAAAGGCAAACAAATTC
AAAAGCTAGCAGAAGACAAGAAATAACTAAGATGAGAGCAGAACTAAGGAGAGAGAGACA
CGAAAAACCCTTCATAAATCAATGAATCCAAGAGCTGTTTTTTTGAAAAGATTAACAAA
ATAGATAGATCACTAGCCAGACTAATGAAGAAGAAAAGAGAGAAGAATTGTATAGACACA
ATAAAAAATGATAAAGGGGAGATCATCACTGATCCCACAGAAATACAAACTACCATCAGA
GAATACTATAGACACCTCTATGCAAATAAACTAGAAAACCTAGAAGAAATGGATAAATTC
CTGGACACATACACCTTCCCAAGACTAAACCAGGAAGAAGTCAAATCCCTGAACAGACCA
ATAACAAGTCCTGAAATTGAGGCAGTAATTAATAGCGTTCCAATGAAAAAAAGCCCAGGA
CCAGATGGATTCACAGCCAAATTCTACAAGAGGTACAAATCAGAGCTGGTACCATTCCTT
CTGAAACTATTCCAAACAACAGAAAAAGAAAGACTCCTCCCTAACTCATTTTATGAGGCT
GGCATCATCCTGATACCAAAACCTGGCAGAGACATACACACAAAAAAGAAAATTTCAGGC
TAATATATCCCTGATTAACACCGACGCAAAAATCCTCAATAAAATACTGGCAAACCAAAT
CCAGCAGCACATCAAAAAGCTTATCCACCACGATCAAGTTGGCTTCATACCTGGCATGCA
AGGCTTGTTCAACATACGAAAATCAATAAATGTAATTCATCACAAAAACAGAACCAATGA
CAAAAACCACATGATTATCTCAATAGATGCAGAAAAGGCCTTCAACAAAATTTAACAGCC
CTTCATGCTAAAAACTCTCAATAAGCTAGGTATCGATGCAATGTATTTTAAAACAATAAG
AGCTATTTATGACAAACCCATACCCAATATCATACTGAATGGGCAAAAGCTGGAAGCATT
CCCTTTAAAAACTGGCACAAGACAAGGATGCCCTCTCTCACCACTCCTATTCAACATAGT
GTTGGAAGTTCTGGCCAGGGCAATCAGGCAAGAGAAAGAAATAGAAGGTATTCAAATAGG
AAGAGAAGAAGTCAAATTGTCTCTGTTTGTGGATGACATCATTGTATATTTAGAAAACCC
CATTGTCTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATA
CAAAATCAATGTGCAAAAATCACAAGCATTTCTATACACTAATAATAGACAAACAGAGAG

FIG. 1D

CCAAATCATGAGTGAACTCCCATTCAAAATACCTAGGAATACAACTTACAAGGGATGTGA
AGGACCTCTTCAAGGAGAACTACAAACCACTGCTAAGGAAATAAAAGAGGATACAAACAA
ATGCAAAAACATTCCATCCTCATGGATAGGAAGAATCAATATCATGACAATGGCCATACT
GCCCAAAATAATTTATAGACTCAATGCTATGTTCATCAAGCTACCACCGAATTTCTTCAC
AGAATTAGTAAAAAACTGGCCAGGCTCAGTGGCTCACGCTTGTAATCCAAGCACTTTGGG
AGGCCAAGGCAGGAGGATCAAGAGGTCAGGAGATTGAGACCATGGTGAAACCCCGTCTCT
ACTAAAAATACAAAAAATTAGCCGGGCGTGGTGGCAGGCGCCTGTAGTCCCAGCTACTTG
GAGAGGCTGAGGCAGGAGAATGGCGTGAACCCAGGAGACGGAGCTTGCAATGAGCCAAGA
TCCTGTCACTGCACTCCAGCCTGAGTGACAGAGCAAGACTCCGTCTCAAAAAACAAACAA
ACAAACAACAAAAAAAAAAAAACTACCTTAAATTTCTTATGGAACTAAAAAAGAGCCCAT
ATAGCCAAAACAATCCTAAGCAAAAAGAACATAGCTGGAGGCATCATGCTACCTAACTTC
AAATTATGCTACAAGGCTACAGTAACCAAAACAGCATGGTATTGGTATGAAAACAGATAT
ATAGACCAATGGAACAGAACAGAGGCCTCAGAAATAACCCCAGACATCTACAACTCTCTG
ATTTTTGACAAACCTGACAAAAACAAGCAATGGGGAAAGGATTTCCTATTTAATAAATGT
TGTTGCGAAAACTGGCTAGCCATATGCAGAAAACTGAAACGGGACTCCTCCCTTACACCT
TATACAAAAATTAACTCAAGATGGATTAAAGACTTAAACGTAAGACCTAAAACCATAAG
AACCCTAGAAGAAAACCTAGGAAATACCATTCAGGCCATAGGCATGGGCAAACACTTCAT
GTCTAAAACATCAAAAGCAATGGCAAGAAAATCCCAAATTGACAAATGGGATCTAATTAA
ACTAAAGAGCTTCTGCACAGCAAAAGAAACTATCATCAGAGTGAACAGGCAACCTATAAA
ATGGGAGAAAATTTTTGCAATCTGTCCATCTGATAAAGGGCTAATATCCAGAATCTACAA
TGAACTCCAACAAATTTACAAGAAAAAAACAACCCCATCAAAAAGTGGGTGAAGGATGTG
AACAGACACCTCTCAAAAGAAGACATTTATGTGGCCAAGAAACATACAAAAAAAAGCTTA
TCATCACTGGTCATTGGAGAAATGCAAATAAAAACCACAGTGAGATACCATCTCACTCCA
GTTAGAATGGCGATCATTAAAAAGTCAGGAAACAACAGATGCTGGAGAGGATGTGGAGAA
ATAGGAACGCTTTTACACTGTTGGTGGGAGTGTAAATTAGTTCAACCATTGTGGAAGACA
GTGTGGTGATTCCTCAAGGATCTAGAACCAGAAATACCATTTGACCTAGCAATCCCATTA
CTGGGCATATACCCAAAGGATTATAAATCATTCTATGATAAACACACATGCACATGTATG
TTTATTGTGGCACTATTAACAATAGCAAAGACTTGGAACCAACCCAGATGTCCATCAATG
ATAGACTAGATTAAGAAAATGTGGCACATATACACCATGAAATACTATGCAGCCATAAAA
AAGGATGAGTTCATGTCCTTTGCAGTGACATGAATGAAGCTGGAAACCATCATTCTCAGC
AAACTATCACAAGATCAGAAAACCAAACACCACATATTCTCACTCATAAGTGGGAGTTGA
ACAATGAGAACACATGGACACAGGGAGGGGAACATCACACACCAGGGCCTGTCAGGCAGT
GGGGGGCTAGGGGAGGGATAACATTAGGAGAAATACATAATGTAGGTGACAGGTTGATGG
GTGCAGCAAACCACCGTGGCACATGTATACCTATGTAACAAACCTGCACGTTCTGCACAT
GTATCCCAGAACTTAAAGTATTAAAAAAAAAAGACCATTTATGAAAACATGACCTTACCA
AAGAACTATATAAGTCACTGGAGACCAATCCTGGAGTGACAGAAATATGTGACCTCTCAG
ATGGAGAATTCAAAATAGCTGTTGTGAGGAAATTCAACAAAATTCAAGATGACATGGCAA
AGGAATTCAGACTTCTATCAGATAAATTCAAAAAAGAAGATGAAATAATTTTTTTAAAAA
TTCATGCAGAAATTTTGGAGCTGAAAAATTCAATTGATATACAAAAGAATGCATCTTACC
AGCAGAATTGATCCTGCAGAAGAAAGAATTAGTAAATTTGAAAACACTCTATTTGAAAAT
ATACAGTCAGAGGAGACAAAAGAAGAAAATTAAAAACAATGAAGCATACCTACAGGATCT
AGAAAATAGCCTCAAAAGCATAAATCTAAGAGTTACTGGCCTTAAAAAGGAAGGAGAGAG
AGAGAGAGAGTGGGATAGGGGTAGAAAGTTTATTCAAAGGGATAACAATAGAGTATCAGT
ATTCAAATACAAGGTTATGGAACACCATTCAGATTTAACCCAAAGAAGACTACCTCAAGA
CATTTAATAACTGAACTCTCATTCAATGGGAAAAGTAAAGTCCTTTCAATAAAGGTGTTG
GGATAATTGGGTATGCAAAAAATGAATTTGGATACCTTTCTTGTGTCATATACATAAAAC
CCCAAAATAGATTAAAGACCTAAGTATAAGAGCTAAAACTATGAAACTCTTAGAAAGAAA
CACAGTAAATTTTTGTGACCTTTGATTAGGCAATGATTTCTTAAATATGATAAAATATGG
TAAAAGCAACAAAAGAAAACATGAATAAATTGGATCTTATCAAAATTTAAAACTTTTTTG
CATCGTAGAATACTATCAAGAGTATGAAAAGAAAACCTACAAAATAGGAGAACATGTTTG
GAAATCATGTATTTGTTAAGGGATTAGTATACAGAATATATATATATATATATATATATA
TATATATATATATATATATATACTCTTACACCTCAACTATAAAGAGACGAATAACCCAAT
CTAAAAAAATAGGCAATAAATAGCTATTAGTTCTCCAAAGTACATACACAAATGACCAAC
AAGTTCATCAAAAGATGCTCATCATCTTTACTCAGGAGGCAAATACAGATTAATATTACA
ATGATATTAGACATGGATTTGTCATATACAGACTTTATTAAGTTAGATTCCCTCTATGCC
TAATTTGTTGAGAGTTTTTATCATGAAGAGATGTTGCATTTTGTCAAATGCCTTTTCTGT
GTCTTTTGAGATGATCATATGGTTTTCGTCCTTTATTTTGCTGATATGATGTACCACATT

FIG. 1E

TATTGATTTGCATTTATTGAATCATCCTTCCACCCCTGGGATAAATCCCACTTGATCATG
GTGTATTATCTTTTTGATGTTTTTTGGATTCACTTTGCTGATATTTTGTTGAGGATTTCT
GCATCTATAATCATTAAGGATATTGGCCTGTAGTTTTCTGTTTTTATGTTGTATTCTAGT
CTGATTTTGGTATCAGGGTAATGCTGTTCTTGTTGAGCGTGTCAGGAAGTCCAAAAGACT
TCTTCTTTAGTGTTTTGGAATAGTTTGAGAATTGTTAGTTTTTTTTTTTTATAAGTTTGG
TAGAATTCAGCAGTAAAGCCATCCAGTTCTGGGCTTTTCTTTGTTAAGAGACTTAAAACA
CACACAACGCACACACAAAATGAAATATCACTTTCCACCCATTATAATTTACAAAGTGGA
AAATAACTCGTGTTGATAAGAATGTGGAAACCTTGAAACCTTCATGCATTGCCAGTGGTA
ATGTGAAAGAATCTTGCCATTGTGGAAAACAATTTGTCAGTTCCTCAAACAGTTCAACAT
AGAGTTACTGTATGAAATAATTCAACTCCCAGGCATGCACCCAAGAGCATTGAAAACATA
AGTACACACAAAAACTTGTACAAGAACAGTCAGATCAGTATTATATATAATTGGCAAAAA
ATGGAAACAATCCAAATATTCATCAACTGCTGAATAGATAAAATGTGGCATATCCATATA
ATTAAATACTATTCAGCCACAAAAATAATAAAGTACGGATAGACACTAAAACATGGAAGA
ACCTTGAAAATATTAAGCTAAGTGAAAGACATAAGACACAAAACCCAACATTTAAAGGAA
ATTTCCAGAATTGTCAGATCCACTGAAGAAGAAACTTGAGTGTTTGCCAGCATGTGGGAG
GAGAGGAAAATCAGTAGTTATGAGGTTTCTGGAATTAGTAGTGCTGATGGTGACACAACA
TTGTGAATATACTATAAACCACTAAATGATACCTCTCAAAATGGTTAAAACATTACTGTT
GTGTTATGTGAATTTTACCTCAATTAGAAAAGAAAAAAATCTTATCAATAACAAAGAGAA
ATTTCCACACAAGGTGGGATCGCTTCCACAGTGCTACTCAATGCAGTTTAGCGATTGCAT
TTGTATTGGAGTAAAAGCATGTCACATTGCTTTTAACATTGGAGTCCAATACATAAACCT
CTTTCACCATAACTATATGGAGTTCATTGTATGTATATTTATTAAAATGGAATTAAGATG
AATTTCACAACACAATGGATCATTTTTTTTTTCATGTGGAAAATCAGAACACATGCCTTA
ATGGTTACATGCCCCACCTGCTGCTCACCTAAAAGTAAATTTCCTCTAACTCAGACAAAT
ATGTTATTTTCAAGGAAAAGAAGCCCAGAGAACTGAGATCCAGAAGAAATAACATGTATT
GAAAGCACACAGAAGTATTTCAATGAACTCAAACCCAAGATTGTAGAAAACTCTCATGTG
CCCCTGGGACTGATGTTTGAAAATACACATATTTTGCTCCTACTCTTTCCTTCCCCAGAT
CCCACCCTTCAGAGCACCCGACGATAATGGATAGTTTCTAGCAGGGTGTCTGGAATGGGC
AAGTACCCCCAAAGTTATAGTTTGTACTGCAAGACTTGAACCCACTCTTTTTCTGCCCTC
TATTATTATTTTTGCATTTTAACCATTTATTATTTTGAAAAGAAAAGAGAATTTTTAGAA
TATGGAAAGAGGAAGTGAATTAATAAAATAGCACACCCTACATAGAGACTGCTAATCCAT
CTCCAGTCTAAAGATTTAGTAATAGGCAAGAATATACATATCCAGGAATTTCCTTGGTGT
TACATAAACAAAGGCGGCACATATGTATATTTTTCACAAAATATTCACTGTTTGAAGAAG
GAATTACTCCCTTCAATTGAGTTCAGGCCTGATCAACAAGTAGTGATTGGCCAACAGCTA
AATGCAAAGTGCATGCTAAGTCTGGGGATACAAAGATGAATGAGAAAACATTTATGCCCT
TAGGAGAAAAACAAATATCTTTATCTCAGAGAATAGAGAAGGAGATTGATTCTCTTTGGG
GGAGATGTCATCCTGAAGAGTATAACAAGTTCCCCTATAATTCTACTTTTCAGTACTGTT
TAAAATACAACTGGATTTTTTTAAATATGTAAAATTTATATAATTTTACAAATGTCTTTG
TTAAGAATTAAAACTATCATTAGTAAAGGACACAGCTGGAAAATTGAAAACATTTTGGTT
CTCTACTGTGGAAACAGAATAGAGTAACAGCAAAAAGCGTATTTCTGGAATTGGACCCTG
ACAACTCTGCTTAAACACTCCACCACTTTCTAGCTATATGACCTTGGGTAAGTTACTTAA
CTTCTTTGTGTGTCAGTTTCTTCATTTGTAAAATTGGAATAATAGATGCTTTTTTTGAGA
CAGTGTCTCATTCTGTTGCCCAGGCTGGAGTGCAGTGGCGTGACCACAGCTCACTGCAGC
CTCAACCTCCTGAGTTCAAGTGATTCTCCAACTTGAGCCTCCCAGATAGCTAGGACCACA
GACACATGCCACCATGCCTGGGTAATTTTTTTTTTAAGTTTTTCATAGAAATAGTGTCTC
ACTAAGTTGCCCAACCTGGAAAATTGGAATAATAATTCATAAAATCTTCCTCCTAGATTT
GTGAAGATCAATTGAGTTAATGTATGTAACGTACTTGGCACAGAGCTTGGCCCATGTAAT
CTCTCAATGAGTGCTAACATTACTTGTCTCACAAAAAGTTACTTACTTCCGTCTGGCACC
AACTCCCTCTCTCACTTCCCACAATCTGGTTACCATTCATTCTTCAGTTCTCAGCTTAAA
CAATGTCTTTTCCATATGGTTTCATTGACGCCACTTTGGGAAAATAGATGTCTCTTCTGC
TTGCATTTTCAGACCTTTTTAGGTGTATACCTTAGGGCATTTGCTTTACTGACCAAAATT
ATTTGCCGGCTACTCTGTGCTTTTCATGACACACTGAATAAGACAGGAAGAGTGTTTATC
TATGCTCAACATAAGATAGGCATATAATGGAAGCTTCGTATATATTTGTTGAATAAAAAA
CATAAGGGGAAAATATCAGATCTAATAATGCAGGACAGGAGGCAAGATGGAACGGAGAGA
ACCTTGTCTGAGAAGAGACATAATTAAAACAGGGCATGGGAGGTAATAGAAAGATTGGAG
GAAAAAGAGACAGAGAGACAGAAATGTTTGTGGTAATTTGTGACAAGTAGCTTTGATTGT
TCATGGCCTAATCTTTTAGGGCATGAGGTTATTTCATTCTCTGTAGCCCACCGAGAGTGC
GTACAGTGACACATGTTATGTAAGTCCCCTTTTCCCTTTTTATAAATGTCTAGACCCCCT

FIG. 1F

GTGATTTGAGACTTTTCTAGAAGAATTTAGCTGAAGACCATATTGTTTTTAAATGTAGT
ATTTGGAGCCTAGAGGTGCCAGATAACTTCCTGCAAAGCTAATGCATTTATTTTGGGAAT
ATATAAGCTCAGTATCATCATTACCAACAGTGCTCAGACTTGATTTTATTTTCATTCCAA
CAGCAAAGGAAAGAAAGCAACTTCTTTCATGCTTCCATGCCACTCTGCATCTCTCTACCT
TCACAGAGTTTCTCAATAATGGCAACATTTCCAGTTCACCAATGGACTGAGAGATCATTG
AGGCTAGACTAGTCTTATTAATCCTTATACCCCAGCTCCTAGCCGAACTCCTGGACACAC
AATAGATACTCAGATACATTTACTGAAATGCATATAGAAAGTTACACCTGCAAAAAAGAT
GATCTCTCACCAGGAATAAGAAAATATAATCTGGGACAGCCCATATATGAGATCTCTAAA
CAACCTACCTATAACCACCAAGAAAAAAAAATACCTGAGTTTGAGATTTATTTTTCCGTC
TCATTTTTAATATATTCCAGTTAGTGAAAGAGCTAAAATAAATGACAAGAAAAATTTAAT
CTAGGTATTTAAACAGAATTATTCTGAATGTTGTGAGCTACATTTCTTTTTTACCTTTTA
TTTATACATAGTATTTGTATATACTTATACAATATATTTATTTTGTATATATAAATATAT
TGTATTTATTTATACATGTAAATGTATAATATATTTATTTATACATAGTATTTATATATA
CATAGTATTTGTATATATTTATAGGGTACATGTAATATTTTGTTACACGCATAGAATGTG
TAATGGTCAAGCCAGAATATTTAGAGTATCCATTACCTTAAGTATTTATTATTTCTCTGT
GCTAGGAGCATGTTAAGTCCTCTCTTTTAGCTATTTTGAAATGTACATTGATGTTAACTA
TCATTAACACAGAGTAATTGATATGTATAGCAAATAATATTTGCAGTAGGATATCACATG
TTTACTTATTTATTTATTTATTTATTTTTATTATACTTTAAGTTCTAGGGTACATGTGCA
CAACGTGCAGGTTTGTTACATATGTATGCATGCGCCATGTTGGTGTGCTGCACCCATTAA
CTCCTCATTTACATTAGGTATATCTCCTAATGCTATCCCTCCCCCCTCCCCCACCCCACG
ACAGGTTCCAGTGTGTGATGTTCCCCTTCCTGTGTCCAGGTGTTCTCATTGTTTAATTCC
CACCTATGAGTGAGAACATACGGTGTTTGGTTTTTTGTCCTTGCGATAGTTTGCTGAGAA
TCATGGTTTCCAGCTTCATCCATGTCTCTGCAAAGGACATGAACTCATCCTTTTTTTGGC
TGCATAGTATTCCATGGTGTATATGTGCCATATTTTCTTAATCCAGTCTATCATTGTTGG
ACATTTGGGTTGGTTCCAAGTCTTTGCTATTGTGAATAGTGCCGCAATAAACATATGTGT
GCATGTGTCTTTATAGCAGCATGATTTATAATCCTTTGTGTATATACCCAGTAATGGGAT
GGCTGGGTCAAATGGTATTTCTAGTTCTAGATCCTTGAGGAATTGCCACACTCTCTTCCA
CAATGATTGAACTAGTTTACACTCCCACCAACAGTGCAAAAGTGTTCCTATTTCTCCACA
TCCTCTCCAGCACCTGTTGTTTCCTGACTTTTTAATGATCGCCATTCTAACTGGAGTGAG
GCACTGGTCTGAAAATATCAATTCATTTAATTCTTTTAACAACCTTAAGGGGATATCATG
GTACAAATTTAGAGCTTTCTTTTGTGTTTGTAAAATGGATTGATTCCTTTTCCCTACATC
CAGCAGAAATATTTGAATTGAAGAGAAGAGTAATACCTAAGAACTAGAAATTCCTTTCTT
ATGTTTCAAAAGATATCAAAAGATCTAAGGAAGATATTCACATCAAAAATGAGTATTATA
ATATTTATTATCTATGGTGCACTTGCAAAAAAGAAAACAAGTAATAATCTGAAGATTTAA
GTGAATATTTTATGACATTGGAGTACCACATATTTAGAAGAAAGCACCAGAGAAATCATA
GATAGAAGGAAATGGAATATTTGTAGGATCAAGATAAATACAGCTTGTCATAAAATAAAG
CAGGTATCAGGATAAAATCTTGAAAATATTTTCATTTCTCGTTATTTATAACTTCAATTT
ACTGTGATGATTAATTGTAGGTGGAAGATTTACGAAGAGAAGACTGAAGTATAGACAAGT
TGAAGTGCCACAAAATGAAAGCTAATGACACTGACTACTTAGGAAATAGCAGACTGGGTC
CATATTTATAGATTGTCAATGACAAGGAATTTGCAGATGTTAATGAATATAGATCCGAAC
TTAAGTTGCAACAACCTTTCCCACTTTGAGATGAATAGTGCATGGAAGAGTAAAATGCAG
ATGTTAATAAATCAGAGGAAGACATCGTGCCAGAGTATAAAGTTGACAGATTTATGCCGA
TGAACTTGAACAAAGCCACAGAAGGCCTACTTGTCAAATTTACTGGTGACAACAGGTCTG
GAGAAATGGCTAATGTTTTGGATAATAGCATTAGAATTTAAGGTCTGTTTAAACTTCAAA
TTAACAGAATGAAATTAATATATGCACATATCAATTGGGTCTTTTGCTTATATATCATCT
CTTAATAGAGCCTTTTTGAACAATCATTTCTAATGTGACCTTTGGGATTTTCTACTCATC
ATCACCTCATCCTGTTTGGTTTGCATTATAGCATCTATCCCTTCCTAACGTTTTCCCTAT
GTATTTGTTAGTTTGTTTTTTTTTAATCTAACTTTACTAGAAAGTAAAATGCATGGAAAC
AGCAACCTGTTTAACTTTGTATCACTAAGAGTGGAAAAATAACCCTCAGGAAATATTTGG
TAAAATAATAAAATGCCCATTGATGCCCTTCTCTTAAAAAGAAATTTAATTAGTGCAGAT
TGGGGAAATACAACAATATTTCTCATAAAATGTGATATCTATACAATAACAGAAGTACTA
TGTCCCAAAAAGTATTCTATAAATAGAAGAAAGAACAGATGGTTTTGCTGCTGATTAATC
CATTTATCTTTCGTAAATCATCTAATTTCCCCAGGAACAGCTTCCTCATCTATTAAAGGG
GGTTAGTAATAGCTAAGCCCTCAGGGGTTTAAAAATGCATATGAAATAATTTTATAAACC
ATAAAGCACAAAACAAATATGAAAAATTATGATTGGAGGAGGGGGTGGGGTAGTTAACTA
AATCTCAGTGTAAACCACCAATGTCTTGTGTGTGTTGAAAAAATAATTACATATAAAAAC
TGGTTGCATCCAAAGAATAATGTACTTTTTGCACTGGCAAGACTCAAACCATATTATTGT

FIG. 1G

TACTTCCTCCCAGTTACATATTTTGCAAGATATTGACAATTGTCTAAAGGAAGACCAAAC
AGATGTAGGTGGGAGCTACTGTCATTTGAACAACATTGAAAAGAAAAATACTAAAAAAGA
AACATGAGGGCATATAAAGGAGCGCTGGGGCTGTGATGTTTATTTTGAATCTGTGAAGCA
TTGTCATGTGGAAGATTTATTCTGTGTAGCACCAAGATGCAAACTAGGAATTAGAGGTAA
AAGTCTCAAAAAGACAAATCGTGGCTTGAGACCTTGGTTTAATGTAAGAAACAGTTTTCT
CACCCTTAGAGCACTCCCATAAGGATGGAAGTAGTGAATTGTGGTGGTCACATTCAAGCT
AGATGGGGACATGTCAGCAATGTTATCAGGAGGCTTCTACTCTGAAGCTGAAGTTCAGAC
AAGATTTCCAGGCTCTTCCCAAGTGCAAGATTGTAATTACTTAAATGCAATATTTTTACC
ATGTTTATTAAGAATAAAAGGATCATGAATTCACATTCTGACAAATGCTAGAATACTTAT
TATTAGAGACAAAACCAGTGCATGAGAGAATGGCAGGTGACATCAGCCCTGAATCAATGG
GAAGAAAGACCCAATGGGATGTGGTATTTACCAGAGAGAGCACTTCTGCTTAGATTGCTA
CATCCTACAGTGAATGTTTAATATCATTGAGTATATTGGTGGTCTGTCATGCTTGACAAC
ATTAACTATGATCATATTTATGACACTTGGCGTCCTTCAAGAATTTGTAGCTCTATTTCA
CATGACACTTAACTATCGCAAATACAAATTCCAGCTAAATAGACCCTTCAGTTTAAAAAC
AGTCTCATTCTCAAATTTTAAGGAGAAAGTGAAGACGGAGATGTCTTAAAGACTCGGCAA
GTACTAAGTTGGCAAATGTCAAATGTTAAAATAAGTTTATATTAAATGTTAAAGTGTTTG
CCTGGAATGACTTTTCCATTGTCCTGCTTGAGAAACACAGAGGCACCTCCTTATTGCTTT
TATATTTGCTTTACAAAGACAAATGTATCAACATGCTCTGTATTAATTGTATGTTGACAT
TTTTGTCATATCCACAGACTGATGCATGTCTGTGCATGGTTTATAATAAGTGCACGTAAA
AATAGAGAAAATAAGTAGAAAAAGAGAGAGATTTAACTCTCACCCCCCACCCCCCAAAAA
AACAGATTAAATTAGTTTTCATTACTTTTTTTTTTTTCTTCAGCTTCAGCTCTCCCTCAG
CGAGGGAGGAGGCTGTGGGCTGCGGACTGAGTGCTGGAATGAGGAGTAATTGAGCTTCAG
CTGAGCCGGACGTAGCTTTCTCCTCCTGGTGTCATTGCTGCAGCCTCCAGTGCCGGGTCC
CTAGTTCCTCAGCTGCCTATCTTCCCGGTGCAACATCGCCTGTAAAGACAGCAAAGCCAC
CGCAGAAGTTGCCCGGCAGAAGACTCCGGAGGCATTGGCTCAGTAACTTTTCACGTCATT
TTCTGCTCGGGAGCCCCTTCTAGCCTCTCCGCGCAGCCTTTCCCACCGCAAATCACCAGT
GCTCATGGGGCAGGCGGAGAGGAGCTTGCAGCATTGAGCGGAACCGGACTTGAGCCCGTG
ATGTCCGGCACCAAATTGGAGGACTCCCCCCCTTGTCGCAACTGGTCATCTGCTTCGGAG
CTGAATGAAACTCAAGAGCCCTTTTTAAACCCCACCGACTATGACGACGAGGAATTCCTG
CGGTACCTGTGGAGGGAATACCTGCACCCGAAAGAATATGAGTGGGTCCTGATCGCCGGG
TACATCATCGTGTTCGTCGTGGCTCTCATTGGGAACGTCCTGGGTGAGTCTCCTCCCGGG
CAGCCCTCCTAGGGGCTATCACCCCCTCTCCGCCCCGGGCTGAGAAGGCTCTAAAGAGAC
CCCTCCCTCCCCCGGGAAGCAAACAAAGAGGTCGCTGCTCTCGGATGGGGTTTTCTAATA
AAATAATAATAATAATAGAAAGTTTTCTGATTTTCCGAACCGGGACCGAGCCCTGGAAAG
GTTATTCCCTGTTTTGCAGGAATAACGGGGAAACCGCGTTTCTTTTTCGAGCACCTAGAT
TACAAGCGCAGGGAGAGGGGCCGCGGCAGGGATCTCCAGGTGGATTTTGTTGAGTGTGTG
TGTGTGTGGGTGGGTAGGTGGGGGAGTCAGTCATCCCTTTGTGTAACGTGGCTGGGTGTT
TCAGGGGGGTTGGGACGAGACAGAGCTTGCAGAATACAAAGCTACATCCCTAAGGAGCAA
GCTCTCTGTGGCTGTGGAAGTCACAAAGCATTTGTGAGCTAGGTGGCATTGCCCCTTTGGC
GAGGAGGTTTAGTCTCCAGTCAAGAGGTGGTAATGAACCAGCAGGGAGTGGAGACGGAGG
CAAAGCAGGGAAGTGCACTCACTCATAGAAGCTGAATTAAACAGGATCCATGCCTGGAGC
AAGAAGGAGGGGCATCGGAGAAAAGTACCACAGAGATCTCAATCATCCATCCATCCATTC
ATTCTTACATCCATTCAGCCAAATATTTTTTTTTTTCAGTCTGCTTGTTGCCAGGCTCAG
GAATTATTCATGTCAACTGTTTGTTGTTGTTTTGTTTTGTTTTGTTTTCTCCAAAGATGA
GACTAAGCTTAATGCTAGGCTATTTGTCCCGGTCTAGGTCTGTATGCAAACACGGGTTTC
CTCGACCCCTCATCCCCCTCCCCCTAAACAATTTCTGAGGGTTGGGGAGGGGGTGAGATG
GCAACATGGTGAGTGCGATGATGGAATGTATTAGGGCAGTTGGGGAATATACCTCCAGAA
AAGGGGCTTTGGAAGGGAGGGATAACTTGAAATAAATTGTGAATGGAAGGAGAGTGTACC
TTGATGAATGAAGAGTAGAAGGCTGGGAGACTTTTCACATGCAGAGGGCAGTGTGGAGGA
AGTCTCTGCTGAAAATGACAGGAGATGGAGGAGGCTAGGAGTTGCTCTTGATTTTCATTT
ATAAAAGAAGAAGAAGGTGAGTGAGGTGAGATAGGCTGGGAGGCTTTGCAGTCAAAAGCA
AAGAACTTGTAGCTGCAATGGGGACTGACAAGGAAATTATCAGGCTTTCAGACTAACCTG
ATTTTTGCCTTCTCTCCCAAGTGTGTTGGTCTGGGTAGAAATCATCCCGAGTAGTCTCTC
ACCAACTCAGCAGGCAGAATAGATGATAGTATGTGAATGACAGGAGTTCTCCAGAGTGTT
GGTAGAATGTTATTTGAGGAGACAAGAAACCTCTGAGAACTTTAGTACATTTTTAAATAT
TATTTTTAGACTGTTTTCCTTTGGTTGATTTAAAAGTAAAAATAAAGGAAATCTTTTTGG
GATACTAACAAAATGAAACAAAAGTGGAAATACACAAGATTAGGATTCTTGTTATAAGCA

FIG. 1H

TAATTCTGTTGATAATAATCCTAATCTTGCTTTCCTTCTTCTTGTTACCCATCCTTAGGA
TTACATCTCTTAAGACACATGGCTACCAGCATAGCAACATTTTACTGCATTATGCCAACA
CTTATTGATAAGTGAATAATCAAAATTGAACATATATTGAGTACCTACTGTGTGCCAGAG
CCCTTCATGTACATTCTCTCCCTTAAATATCAAAATAACCCACATTAGCCAGAAGAAGAA
ACAAGACTTAGAGAAATAAAATGACGTATTAAGGGACATAATTTAAATTCAGTTCCATTT
TTTCTGACCTCAGATCCAGAATTCTCCATTGTTATTCCACTCTAGAGCTAAAAAGCATAT
AGAGAATAGATTCTCTGCTCCTGATTGTCTGCAAGTTTATTAGATGTGTTCCTGTTCTCC
TCTGCATCAACGCCCACTGCCAATAAAGTACAATGAGGGATTAATGGCACTGTCATTCTC
TTCACCAAAAAACCTTTCCAGAGAAGCAGTAATTTTTTTATGAATAGCTATCAATAGTAAC
TATTTGCCTTCCTTATTTTAATTTTCGGCTGAATCTTTGTGGTAAAATGTGCTCTTCTTT
GTTGTTATTGCATTTTTACCTTGCATAGACCTTGTAGTGAATAGTCTCCATATCCTAATT
GCATAGTTTAGGGATACATGTTTGCTAGCCTGGGGAGTTTTAGTTTCAAGAAGGAAACAC
CTCTACAGTAAGGCTACTTGTTTCATAATGTCAAGGAAGATAGCACTGTCCACAGCCCCA
AGTGCTGAAATGGCCAATTCCATTCAGCCTAAAAAGAAAGATTTACTCAAAGCACTCTGC
CTTAAAGAAACTGACAGCTATTTTCCTCAGGACTGAATAACACTGAAATCCTCTCTGGTT
GAACTGAAATGCATTCTTTTCTGACATACTGCCTGAAAGTTGATGAGGTTTAGGTTTGAC
ATTTAAACAAACGAGTAGTGTCGTTACTCACAGACAACTTCCTGCTCTTTGATGTCACTG
TCAAATTTGCAAAATGAATTAGATTGAGAATTGCTTCTTTGCCCCTCTGGTATAAGTAAT
TTTGCACATAGAGTGGTAGGACAGGATGTCACATGATTTATGCAAAATAAAGATGCAATA
TTAAGTATGAAGGTAAAATACCACAGTGTAGGCAGCAGATGTAATCACTGAGCCTTCAGG
TCCAGTCACCATTTGTACTTTCATATAACTGCTTGGAAAATCTCAACCTTTTTGGGCTTA
CAAATATAATGCCATCAGTTAGAAGTCATCTTCTCCACAATGTCCTTTCATGAAGTGATG
TAATAGGATATGCTGTGGGTAGCATAACAAAGTCTTGATTGTCCTCATCTCTTTTTCTTC
TCCCCATAGTCCCTCTTTATCACTATGCCACCTCTCCACTCTCATATACTCCTCCCAAAG
ATGGAAAGCAGTTTCCTGGGGGAGTAAAGTTTTAAATAGAATGTTATGAGTATTTACATT
CAATGAAAAGCTGTAAGCATGTTTAATGTGAAATTTTAAGTTCTAAGGAAGGAGCATAGG
GTAAGGTTCTTTTTGGAAGGAGTATCTTTTCAGTATCTTCAGAATAATGCCACCTATAAC
CTATTCCTAACTATGTCTTCTACTACAGCTAAGTAGATGTATCAACTTATTCAATTGGTA
TATTGTGAGCATTATCATTTTTTTAAATTAGTGTGTATATCAGGGGAGCCTCTGGGGAAA
TGTAAAGAAATGTGACTGATGTTAATTTTTACTCCTGATTCCTTGAATGACAATTGTAGG
GAGAAATGTGTTCTAGTCAGTTTAAACATTAAGTACCTAGGGAAAATGATCAATTTTCTG
CTTCTCATATCTGCATTCAAAGATATCATATGTTTCATCTGGTATGCTTCTGTCATATCT
GTTGTTGTCTCCATATGGAAAATAGGAAAACATCAGTCTAGCTATGCTTCTTGCTTCTTG
TGTGCCATTAGCAAGTTATTGAACTATCCAAGTCAATTTTTTTATAATTACAAATTAAAG
ATCGATAATGACTGCATTATAGAAATAGTATCAGGATATAATGTACGTATACCCTCTATA
AAGACATATAAAGGGACACAGGCATATACATATTTTTCTTGACACATAGACACTAATTAA
TGTCAATTTTTTATCCCTTAATTTTCATGACTGAACTTTTTGTGATGTGGTGTATAGCCAG
CTTCTGCCTTCATGGGCCAGTCTGTATCTCTGTAGCTCTTTATGGCCTCTGCCCCAGCCT
TTTCCTTAATTGCATATTTTCCTAAAAGGTGTGAATAAAATGGTGTTGGCACACATTACT
CTCCTTTTCCACACTAGCTCCACCCACCCATCTCCTTCATACTGATTGCTTAACATTGCC
TTCTTGCCTTTAAATGAAAGCCATTCCTAACTATTGGAATAGTTTGCTTTCTCTCTCAAC
TTAAATTTGCCTGTGCTGGGTCCCATTCATTTAGAGTTTTTGAGTTGTTAATAGGTTGTT
GATAGGCAGGTCTATCACTACTAGTGTTTTAAATAACACACACATTGGTAATATGTTGAT
TTAACTCATACATTGTTAAAATACATTGTGAAGTATTCATAGTTAAAATAAATTATCCAT
TAAGTAATTTACCTAATAACAGTTTACCCAAGTTAGGTGTGTGGAATGGGGAAATATTTG
TAATAAGTTTGCTTCCTACAGAGTTAGTCTTGTGTCAGATATGTAAGTGGTAGAATTGCA
AGTTCATGTTACTCCTAAGCCTAGAGACATTTATTTTCTGCTTCTCCGAATGCCCATTTT
AGTTTCATGGGTGTTTGTAAACCCATCCTTACCTACACAGGAAGCAAAAAGGGGTTATTT
CTAAACCCCTTTTAGATATAGAAATAATACATCACTCATCTCGGCCAAGACTCAATAGAA
TCATGAATAGTGACTGTAAAAGGTAATATTAACTATTAGGCTTTAAACCTATTGTGCATT
TTAGTTTTAAAATGCAAACATGCTAATCTGAATAAGAATTAATCTGATGCCTCTACATTT
TTGCTAAAATCATAACTGTTTAGTCTTACTTAGTAAAATAAATTATATCTTTGACTTAAA
ATCCCAATGATAACTTTTAAGATGGCTATTTCATAGATAACAGCAACATTTATCATGGAC
AGACAATAATGAGAATAACATGTGCAACTGATAATTTAAATGCAATGAGTTATTTCTGTA
TTTGAAAAAATATATTTGGGAAATGGGATAATTAAAAAATACCAGTTTTCAAGAGACCAA
ATCTAAAACTCAAACATAAACACAATGCTCCAGTTTTTAGAAAACTGTCTTGATTGTAGT
AGTGCCTACATACTAAATTGTATCATATGATTTATATTAATTTTCCTTATTTTGTATTTT

FIG. 1I

AGATTATATTTGAAAATTTTCATGTACTGCAGCTATGTTAGCATCTCAAAGTCTCCATAT
TCTCACTCCGCTCCGAAACATCCACTGCTGATGTTATTTAACTAGTGAAAGAAGATCCTT
CCATGTTTCTTCTTATAGCATTCTGACATCTTCTCCACCCTAAGGAATGCTGGCTTTATT
AAGTATGTTTCAGTCAATGACATGTGATTGGTGAAGCTGACGGTATTTGTCTTCAGTTCC
TTTTTTCCCTGCAAAGGAAATTTGTTGAATATTTATTGGGTACTATATGCCAGGTACTAT
ATGTCAGGCTCCACTTACATATACTCTATTGATGCCTTACAACAAACTTATAATGAGAAG
ATTAATAGGTTTTACAAATAAGAAAAATGAATTCAAAGAGCAATGCTAACTTACTCAAAA
GTTTAGTCAGGCAGTAAATAGCAGCACTAGGTTTCAAATATGGATTTAACAAATTCCATG
GTCCATGCTTATTCCATTACTTCATCCTGCCTCTTTCCTTAGCTTCTAACCCTGACTGGA
GATGCATAGGCAAAAAGAGGAAGGAAGAGATACTTAGATGTGCCCTCTAGACAATTTACA
GAGTTGTTTGGGCATGTTGCCATGCTGTTTTTCTGATAGACTACAGTTCTTCAGCTCTGA
GGATGAGCTCATTTGATAAGCCAATCAAGGTCGGGCTAGGGTTACTTTACAAGAGAAAAT
TTCAAGGTAAAATAGGTGCTGCCAAAAATGCTTTTACCTGTTCAGGGGGTTGACTCACTG
GAAAAAAAATGTTAGATAATTGTGGCCAAGGATTATTTTGTTATTGAAAGTGCTATTTTT
AGACACAATTTGAGCCTGAGAGCCTAAACACTTAACACTTCACATAATCTACAGATATTT
GTTTATTTTTCTTTTTGTCATGCATTGCCAAATAAATAGTATTTATTTAAACAAATCATG
TTGCTATTGATTTTATTAAATAGATGAACTTTTTTTAATTTTTTTTTTTTGAGATGGAGT
CTTGCTCTGTCACCCAGACTGGGGTGCAGTGGCACAATCTCGTCTCACTGCTGCCTCCAC
CTCCTGGCTTCAAGCTATTTTCCTGCCTCAGCCTCCCCAGTAGCTGGGATTACAGGCACA
TGCCACCATGCCCAGCTATTTTTTTTTTTTTTTGTATTTTTAGTAGAGATGGGTTTCACC
ATCTTGGCTAGGCTGGTCTTGAACTCTTGCCCTTGTTATCTACCCACCTCAGCCTCCCAA
AATGCTGGGATTGCAGGCATGAGCCACTGTGCCTGACGTGAACAGGTCAATTTCTATATC
ACCGGACAGTGTTCCTGGATCAGAATAATATATTATATGTATGAAGAATCATTACCTATT
ACATCAGACATGAAATGACCTTTAGATACTGACTTTGAAAGAGTTTGAGATGCTATTGGA
TGAAACACATGACCCATATGACCAGTCTTTTGAATTGCTGACTCTGAGTATAAAATGTTT
TCATTTCACCTTTGTTCACAATGAGAAGTGATCTCTTAACCAAGTAAATGAATTAAATCG
ATATTTAAAATAACATTAAATTTCTTGCCAGAAAAACTGTTCTTTCATAAACAAAAAACA
AATTGCTCAAAATAAATGACTATATCTTTATTTCTAAAAAATGTTTAGAGATTATTATTA
TTGGGTCTTTACAAGTAATTTGCCTTCAATACTAAACACATGAGAACAATGTTTAATATT
TATATAGTATTTTACTCTTCAGAAGATATTTGTCCATATTCTCTCTCAGTTATTCTTCAC
AACAACATTATGAGGTAGGTCTTTTTTAATGAAAAAAAACTCAAGTGCTTGAAGTGATTT
AAAATCACTGTGGAAGAAAAGCATGGGCATACAGAAAAGCCAAGTGGTTGTGTGTCAGCT
TGGGAAAAGCTTGCAAATTTCCTGTATTTCAAGAGGCCAGGATGAGGTGTGTAATTATCT
TTTACTGGTCTTCAGCTATCCTGTCTTTGATATGTGATTGTGTCAAAACTATGAGGAAAA
ACTCACATTAACAAACTTCATAAACTTGTTAAACATAAAATAATAATTTCGATGTTTTAA
TTTACAGTAAGAGTTTATTCTTACAAGTCCTTAAATACCCAAAGTTCTTTCAGTTATCAT
AGTCTTTTTCAGTAGACAGAAATCCATGTGGACTGTTATTGTTCTGAATAGCTAGGCTAT
GCCATAGTAGCAAACAAACCCTGAATTTTCATTGGCTTAGTATCACGAAAGTTTATTTCT
TGCTCATTTAACATCTGAGGTGGGTTGGAGAGTCTCCTTCATCCAATGACTCACAGTTCA
GGCAGCCTCCACATTTTGTGCACTATCCCTAAAAGGTGGACTCTGTGGTAATCAGTTTCC
AATATGGCTTCCAATGACCGCCCCCGGGCCCCGGCCCCACTTCCTGATAGTCACATCATC
GTGTAGTCCCTTTGCATATTATGCCAGAATTGGTCTGGGTGACCAACAGCTCATAGCAGC
AGTGAAACGATGTCACTTTCAAGATTACATAACAGGAGCTTACAGCTTCTGGCTCAAGTA
CCCACTTTCTCTCTAGCTCTTGGATCTCTTCTTCTGGAGGAAGTAAGCTGCCTTGTGGTG
AGCAGCTGTTGGCTGGAGTTAAAATCTCCAGCCAGCAGCCAGAGAGGAAATACGGTCTGT
TAACAACCTCATGTGTGAGCTTGGAAGCAAATCCTTCAGACCAGGTTGAGTCTTGAGGTG
ACTACAACAGCCACTACCCCAACCCACCCCCAGCTTCAGTGCAACTTAGTAACAGACACT
GAGTCAGAACTATTCAGCTAAGCTTCTTGCAGATTCCTGACCATTCAGAAGCTATGTCAT
AATAAATTTTTGTTGTTTGACTTCAGTTTCGGGATAAGTTGTTGCACAGCCTCTAAAGTT
GTGAACTAGAAGAAGTATACTGGCTCTTAACCACCTTTGCCAAAAATTAACACTTGTCAG
TCATGGTCATATTCATTTGGTCCAAATCAATCATATCGTATCAACCTAACTACAAAGGGG
ATTGGGAGATGGTGATGTCTCTGTCACAGAATCTATATAATAGTTAAAAGTATTTTTAAC
TTGCATAGACTCAGAACAAGATAATTTGGAGGAATTCAATGCTTAATGGCATACCACTAA
GATAAGCTGATAGATATATCGTTGCGATTTGGGTCTCTGACAATAGAGGCAATTGATAAT
ATTAAGAGACTATGTGCCAATTATTGTGCTTGGATTGAGGGTACAAAGGTAATAGAATCC
AAGGAACCTGCACTCTTTTTGAAAGATAGACACATAAACACATACTTTTAAAATAACGTG
GTAAGTGCTACTATGACAGATGGTTGCACAGAATGTAGTGGAAGTATTTGAGAAGGACAC

FIG. 1J

TTAGCTCTGCTGGGGGATTAGAGAGAGATACAGGAGGAGATGACACCTAAACTGAGTTTT
AATAGATGAATTCAAGTTACCCAGGTGAAGAAAATTGGGTAAGGATGTTCTAAGCAGAGG
AAACAACATAAGCAAAATCAAAGAGGCGTGAAATAGAATGAGCTATGAAGAAAGTGTTAG
GCAATTGGGTAAGTCCAATGTAAGTGCAGATGAGGAGAGTCTGGAAATGAGGCTGAAGCA
GTAAATAAGGATTGGCCATAAAAGACCTTGTGTACAATTCTTAAGATCTAGGCTTTGACA
CTGTTGTTTAGGGGGAGCTGTTAAAGGATTTTAAATTAGAGTACCATCATTGGTTTGCAT
TTTCCATGAGAGCATTTTGAGGAAAATGCAGAGAATAAATACATGAGGGGAAAGACTAGT
GAAGGTTTTCACACTGGGGTTTGCATCCTGTTTTGGCAATAAGCTTGTTTTAATGAAAAC
AAACAAACAAACTGACAATAAAGAACATAATCCAAATTCTCCAGATAATTACTTCCAGGA
GGCTTTCTACGTGCTGCATACAAAACAAAGAAAGAAAAACATAAAGTGAGAAAACGAAGG
AAAAACAAGGAAAGAAGAGAAAGAAAGAATACATATTGGAAAAACTGTTGCTGTTTTTGT
TTTGCTGAATATTTAAATTTGAGAAGCAATTTCTCTTTTTCTTTTTACTTTTTTTTTTGA
GATAAAGTCTCACTCTGTTGCCCAGGCTGGAGTGCAGTGGCGCCATTTCAGCTCACTGCA
ACCTCCGCCTTCCAGGTCCCAGTGATTCTCCTGCCTCAGCCTCCCCAGTAGCTGGGACTT
CAGACATGCACCATCACGAGCAGCTAATTTTTTGAATTCTTAGTAGAGATGGGATTTCAC
CGTGTTGCTCAGACTGATCTTTAACTCCTGAGCACAGGCAATCCGCCCACCTTGGCCTCC
CAAAGTGCTAGGATTACAGGCGAGAGCCACTGCACCCAGGCGCAGGTTTTCTTTATGATG
TTTTAATTATATCTTTCTTGGAACATATATGTATGAATCTTGCATGCCATAGGTCTATTA
ATATTTTCCAATATTCTACATGGTTTTTTACTAAAATCATTTTTATGATTAGTTACTGAC
TGAGGTTTCAATGCATCACTGTACTCCTAGCTATCTCTCATTTTAGCTTTTACATCACAT
TTTGGCCTCACACTGAAACACAAAATATTAAAAATTTGAGATCTAATAAACAATTTTCAC
ATTTTCCAACTAAATCCCCACTTCTTTCTAAATTTTCTACAACTTTCTAAACATTCTCAC
TTGAAAATTTATTTTAAATGACATGTATTTTATTCAAACAATCAATGAAGATGCTACATT
GACCCCAAGTGAGCCCTTAGGGAATTTCCGTGAATATTTCCCTACAGGTTGGCATGGTAA
CACACTTCACAATTTCTAAATCTGTGGATAGTTTAGAAGCTTTTATTTGCTGTTCCTAGT
TCACAATGGAAATACAACAATGATTAAAAATTATAATATCCTTTTGTAGATTCTTAGCTT
TTATTCCTACTCAGTGACTCTAAAATGAATTTATAAGGCCCATGGTTTATAACCATGTGA
GGCCTTGATTTTGTCACTACATTGCTAGAAATGGGGTCAGAAGGCCACCAGCTTTAATAA
TTTAATTCATCAATTCGGAATGAATTTGATGAGTCAACCACTTTGGTAGAGAACCATATT
GCTCATAAATACTGTTTTGAAGGCAATTCGTCTTTCATAAAATGTGAAGATTGTGCTGAT
CTTTCTGGGCAGGGTTATGGAGGTGTGATTAAATGCTTAAGAAACCATTTTGTTATTATA
TTAAACCGAATCAACTTTTTATTATTAAAAATAGATAAAAACTTAGCATCCTCAATTATA
ATACTTTATACAAAAGTTTCCCAATTTTATATAGACTGAAGATAAAAATACATTAACAAA
TCTTACCAGCTGGTTCAGGAAAATAACTTCATAATTATTGAGACATTTATGTGTTTGGGC
TTGATTTATACTTTGGACACAGGAAAACCTAGAGAGATCTGGTTCTTTGAAATCATCAGA
GATGGTGATGGTGACTCAGAGATTCCTGAAAATCAGTAAGATTACCCTAGTTTATAGACG
TATGTGTTATTTTTTCCCCCAGGCATAATGAACTTTATAACTTGTCATTGACAAGAAGCC
AAATCATCTTAGAGAAAAGGGGGAGAATAAAAATTTAAGAACTTAAAAACACATAAATAA
AAACATGTACATACCTCACACATGTGTACACACACAGTTTGGGGATTGGATGATATGAAT
AATATAATTAATACACCCTAATTTTTCATGCAGGATTAAGAAAGTATCTTCCAAACATTA
AAAATGCTGAAAACTGGACATAAGGCCTTGAGTTTCCCAAATTCAGGACATATTTTCAAC
TATCCCCTGAGTAAATGAACTATAACATTTACAGAAGTAAAAATGATAAATACACTAAAG
ATGAATAAGTCCTTGAATTAACAGCCAAACAAGAAGGCGCATCCTTTGGATGATTGATCA
CTGTAGCATGATTTCTTTTCCTTGAATAGACAATATTCCTTGACAATCTTTCTGTAAACA
GAATACAATGTTTCCCTAAGCAATATATGCGTGCTCTAGAGTTTTCACAATTTCTGATCC
TCCTATGACTGGCTCCTGCTCAGCTCACACTGCACTTTCATGGAAGTTCTCTTAGAATGC
CAGCTTTGAATCACTGCTCCCTCATGTGCTGTGTGTGATAGCATCCCATTTTAGTTTTGT
CATAGAATTGATTACCATTTCAAATTGAATTGTTAATTTATTGTTCATTTTTCTGTTGTC
TCCCTTAAGTAAAAGGTAAGCTGCATGAGAATAGTTTCATTTTTTTTCCTGTTTGCCAAT
GTATCCTCAGTGCCGAGAACAGGTTCAGGAATACAGAATTTTTAGTTAGCAAATGAATTA
AAGTGTAAGACTTCCAGCAGGAGGAATTTTTTACATATAAGTACATTTTTTAAATTAAGC
ATTGCAGGCTTTAAATTTCTTCTATATAAATATTTAAAATAAAGCTTCAATAATTTGAAT
TGCTTTTGTGATTATTTTGTTTTATACCTTGAGTAACTTATACATCAACTATTTTGTAGT
TATTCTAGTAATGATTATGAAAGACCATTTGAAAATCTTTCCCCAGCACTGAGATCTCCT
TGACATGACTAAGTGATTTATACTATGCAATTATATTGCTCTTCTCAAGAAAAGCAAAAT
GAAATTTACAAATTTGGTAGCTTTTTGTTCTTTTGTTTTCTCAAGTAAGATACACCAAGA
TTTCTTTAAATGATACGCTATATTTCTGCAATAACTGAGAAGAACATGTAATGTGCAAAA

FIG. 1K

CTCTTAAACTCTTTTGTTTCAAAATAATTCTTGGTTGTTTTATAAAAGTCTAAGCAAA
TACTTAATGAACTGTGTCCCAAATGAGGTGAAACAGCTGTGACAGAATGTTACTATGACT
CTGTACTTTCTATAATAAAAAGGGACAGACATATCCTCACCTGAGCCTTGGGATGTTTCA
GGCATGCCCATAGAGCCTAAGCTTTAGGAATCCTCTGTCATTCTTTTCCATTGCCAGTGA
CTTGTGCCAATTCTAGGGTTCTGGACTGTGCAAACAATGGAAAAAATAATAACACTTTCA
GGTGGCGCACAAAACCAATGTTCATAGTAGATGGATAGTTCTAGACACTTTATTTAATAG
AGAATAGGAGAAACACTAATCCCATCTAATTCTGCCTTCAAACTCCTAAAATATTCATCA
TTATGAATTAAAAAAAAAAATCAAAGTGTAACCTCACCCAGAGAAAGAAGACATTGGGGC
CAGGTCTGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGATGCTGAGGCGGGTGGATCA
TGAGTTCAAGAGATCGAGACCATCCTGGCCAACATGGTAAAACCCCATCTCTACTAAAAA
ACAAACAACAAAAAAATTAGCTGGGCTTGGTGGCATGCGCCTGTAGTCCCAGCTACTTGG
GAAGCTGAGGCAGGAGAATCACTTCAACCCGGGAGACGGAGGTTGCAGTGAGCCAAGATG
AAGCCACTGCACTCTGGCCTGGTGACAGAGTGAGACTCCGTCTCAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAGGAAAACGAAAAGAAAAGAAAGCAGATATTGGTAATTCT
AGCAGATCCTGGAACAACTGAACCAAATTTATTAATATGTATTATTACTGAAAATCAGTA
ATGAACAAAATTTACAGAATGGGCTTCTTGGAGTTGTTACATTTCCCTTATTACATAACT
CTTCAATAAAAGTGTTTGTCATACCTATTTTAGTTAATTCTACAACAACTAGTGTGATAG
GGCTATTATTTGATCTTTTTTTTTTTTTTTTTTTTTACAGGTAGTGACATTCAGTATTA
GACAGCTGCTATTGTGTTAGTTGTCTGAATACCTTTACATATTATCAACTGGCCTTTTCA
TTCCTGAGTTGTGAGTAAATGCTCTGTCTCCCAGACTGGAGTGCAGTGGCGCAATCTCGC
CTCAGTGCAAGCTCCGCCTCCCGGGTTCACACCATTCTCCAGCCTCAGCCTCCCGAGTAG
CTGGGACTACAGGCGCACGCCACCATGCCCGGCTAATTTTTTTTTCTTGTATTTTTAGTA
GAGACAGGTTTTCACCATGTTAGCCAGGATGGTCTTGATCTCCTGACCTCGTGATCCACC
CGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCACACCCGGCCATAAAT
GCAGTCTTGTGTTCCCCACTTCCATTCCTCCTTTGACAGTACAGCTATGCTAGTCTGCGT
AGCAAATTGAAAAAATATGACCTGTGGGATTTAAACAAAACACAGTGTCATACACATTTT
CTGGTAAACTTAACCAAAAGGGACTTGGGTTCCATAACTAATCACCAATGCCTCAGTGAT
CTGTAACTCCTTGTAGGTACCTGATCACAGTTACTAAAGGGAAAGAGGAGCGAGGAATAC
AAGAGCAAAGTCAAGCCAGACATAGATTTTATCTCTTTGTAAACAGGAGTTCAGAAGACC
GCTCTGAATGCTGAGTTAGCATCAGCAATAATAGAAATATATGCAGATTGTTGATTTGAA
GTCATGCAAAGATATCTTTTTCATCCAAATGGAGGCAAAAGCATCATAGAGCACCAGAGG
GCTAAATCCAACTGTAGCAGCAAAAGGTACACAGAAAAAATAAAGCATCCTGAACCAACGC
ACTGACTTTCTAGGGCTTATCTAATTTGGAGCTATTTCCTTTTCTTATTTCATTCAGCAA
ATATTTATTGAACACCCACAATGTGTAATCTGTTCTATTACATTCTGTGGAGGAAATACA
GAAGTGAATGAGGCATGGTTCTTACCTACAAGGAATTTCTAATCTTGTGGGGGAGACTAA
CATGTAAACAATAAACTATAGTATGAGGATTACTGAAGAGGCATATGCTAAGTCTCAGAA
CATTGAAATATAAGAGTTGGGTTTGACATGGGGAAAGAAATACCTTCTTCACTGAGGAGG
TAGCATTTTGAGTTATTGTTGACATGTGAATACGATTTTGAAAAGTTCCAAAGAATGAAA
AATTCCACCTACATTGGTGAAGTACTAAGATTAAATGCATGATAGCTTGAAGACACAAAA
ATAATTATTTATAAACCATTCCAAAAATCATTCAGGGAATTCCAATAATACACAAGTTTT
TAAACACATTTCTGGGTAATTTTGAGTAATAAGGTCTTAATCTCCTCTACTGCTTTCAAT
TGTTTTTGTGGCCTTCTTTATTTTGTGGGTATCTGGCCCAGTCTTGTCTGTAGTGTATTA
TGGTGGATTGGATTAAACATGTTTTGCAATCTCTGGAGTGATTTTAAAATGACTTGTGTT
ATATCAGAGTTTCCTAAAGGGAGATTAATTTGGCTTAATGGTAAGAACGGATTAAAGTTA
TGAGATACCAGACACTGGGAAAACAGTTAGAAGCCTGTTGAGACTCTTCAGGGCAGTTGT
TGTGAGAATGAAGTTAAGACAATGGGATAGAATATGAAAAAAAATGAAACAAACATGAGA
GGCAGTCTGAAGATGGAAGTTGGCAACTCATCAAATGTGAGAAATTTATAGGAACAGAAA
AGAACCTGCTGATTAATATAAATTTTCTGCCAAAGAAAGTACAGTGGCTCTCCTCAGCAA
ACTAACATGGGAACATAAAACTAAACACTGCATGTTCTCACTTATAAGCAGAGCTGAACA
ATGGGAACACATGGACACAGGGAGTGGGACATCACACACTGGGGCCTGTTGTCGGGACTA
TGGGAGGGAGACCATCAGGATAAATAGCTAAAGCATGTGGGACTTAATACCTAGGTGATG
GGTTGATAGGTGTAGCAAACTATGATGACACACGTTTACCTATGTAACAAACCTGCACGT
CCTGCACATGTATCCCAGAACTTAAAATAAAATTAAATTATTAAAAAAAGAAAAAGACAG
TGCTTGTCTTATTCGTTTTTTTCTTAAAATGGGAAATATGTAATATATATCAACTGTAGT
GTATAGAAGGGTCATGATGAATTGGACAAAGATACGTGGAGTTTGAATTGCTAGAGGAGT
ACCCACGTGCAGTTGTCCAGCAGAAATCAGGGCTTGTTCCCCAACATGCTATTCACAATC
AGTCTACTACTCTCAGGTATTTGTTTTTCTGTGTGGCTATGCAAGCAATAGATACAGTTT

FIG. 1L

ATGTGAAAATGTTTTAGAAAATGTCTTCTGGAATAATTAAAAGCATACAAGGGAATGTAA
ATCTCTTAATGTGACAAGACCTTTTTGCCACAATAAACAAATTCATTAGTTCAAAAAATA
TTTATTGTGTGCCTATTGCAGCAAACAAAACAGACGAAGCTCCTTCTTGTAGGGAACTTA
TACTCTAGTGATATTTAGTATATATTTTGACAATTGAACCAACAGGATTTGCTGACGGAT
TGCCTTATGGGTATAAGAGAAAGAGAGGAGTCCACACTTTCATGCCAGGTAGGTTGATGG
AGGTGCCATTTACTGAGATACAGGGCCGTAGAGGAGGAGTGTGTTTGCAGCAGGGAAGGA
GAAGACTCAAAATTTGGTTTTGATCATACTAAATTTGATATAGTACAGGTAAGTGTATGG
TGGCCATTAGAACATGAAGGTAAGAGTTTAGATAAGGAGACAGGTATGGTGAAATACATC
CAATATTTATAACCAATATTATCTTTTGTGTCTGTACCTTTTTATACATTCCCCATATAT
ATCAAAGACTATAGAAGGGACTGGATAGTGAATAAGTGATTATACATAAATTCTTTTTTA
CAGATTATTTTGCTCTTGATTTCTCCTATGTAAATCATCACAGCTACATTTTTTAAAATC
TTAAAAAGGATTACTTTGAACAATGCATTTAAACATCCAGAAAACAAAAACAGGAGTGCA
TGGTAAAAATTCTGATTTCAGAACGTATGCCTGACTTATCAAGTCAGAATTTCAGGGAGT
GAAGACCCTGGAATCTACACTTTAAATAGAGCCTCAGTTCACCAAGTATGAGAAGTCCTG
TAACAGGGAAAAGTAACCTCCTGTTATATTTGATGGAGGCCAATTGACAAGCCAAGTAGT
TTTCCATTTGACAAAAATTCTATTGTACCAATGAAGAGCTATCAGAGGGGAGTAGATTAA
AACACCTCCCTTGAAATGGAATTTGGCAAGAAAGCAAGAAATTACAGCAAAAAGACCAAT
AAGAGGAATTAGGGGCAATGAAGGAAGGAGCAAAGATGTGGGAACCCAAAAAGTTTTCCT
AGTAACAACTTTGAAATTATATTTTTAGTATATTAAATTTAAAGTAGAGTTATTAGTGCA
TACATTGGTGTAATTTATTATTATATTAAGCCAACAATATACTTTTAAACTTATACAACT
TTGCAAAAAAGTACAAATCAGAAGTCTGGGCTAAGTAGAATGCATAATAGAATCAGTAGT
GCAAAATATTGTTCTATATTTTCTAGCTTATGATTTTCTATATAAAGTCAGTCTTTCAGG
ATTAAAATGAATGTCACTTCTTTTTACCATGTGTCCTTTAAATTATTAAAATCTATACAC
ATATTGCTATACATAGTAAATATAGTTAGTCAATTATGTCATGGAAAGAATTGAAGGGTT
GTTATAAATTTAAAGGTGTTTCACTATACAAAAACATTGTGAAATACTGGTGCTGATTTA
GTTCTAGTATCTCTGATATATTAAATCATAAATGTCAGGAGTTATTGGTCACAAAATAAA
CACCAGAATTATATGACAGTCTAAAAACAAAAACAAAAAACTTCAGCAACAATATTGAAG
ATATGGAAGTGCCAGAAGAATAAGGATTAAGACAATGAATAAAAATCTCTTCCAAGGACT
GGTCTACACTAAGAGTTTAGAAATGCATTTTTTTTTCACAGAAATATCCTTAATCCTCTA
TATAGAAATGAGAAGAAAACATAAGACTTTAGCAAGCTCCATCTAATCCATTTGCAGACA
TATGGTTACCTATCTTTTCTTCAATATATTGGAGTTTGCAAATATTCTACCTTCAAAGAA
TAGGTGTTACCAAAACATTGTCTGCAAGATTTCTAAGATTTGAAATATATTTGCTATAGT
AGGTTAGAGATGAGACATTTTTACTTTAAATTGCAATAATTCAGACTTAAAATATAAAAT
GTGTAAGTCTAAATTTTTTTTCTATTCATTGCAAATATATCTTATATATACATAAAATCC
TGTGTATACTCATATGAACTTTAAGGAAATATCAGAGGCATCAGTAATAGATAACTTGCA
TCTCTTTTACATTCAGTTCAAGCTACTCAAATTTTAATCTTTTGTTTTCATTCCAACAAA
AAAAATTAGGATCTGCCTTGGCTTTTGCTAAGAAAGTAATTATTGGCTGGACATGGTGGC
TCACATCTGTAATCCCAGTACTTTGGGAAGCTGAGGTGGACAGATTGCTTGAGCTCAGGA
GTTCAAGACTATCCTGGGTAACATGGTGAGAACCTTTCTCTAAACACACACACACGCGCA
CGCGCGCACACACACACACACACACACACACACACAAATTAGCTGGGAATGATTACACGC
CTGTGGTCCCAGATACTTGGGAGGCTGAGGTGGGAAAATCACCTGAGCCCAGGAAGTCGA
GGCTACAGTGAGCCGTGATTCCACCACTGCACTGCAGCCTGGGTGACAAAAAGAAAGTCA
TTATCTTCAACACTGTGCATACACACTTTTCTGCATCTAGATCCCAAATTTTTGTTTTGT
ATTTACATAGAACATTGATAAGTAAGGTAAGTATTAATTGATAAAACATTTCAAACTCAT
TTTTCACTAAATCCAATGGCCTTCCTCTTTTGCATGAAGTCTCTAAGAATCATGTTAATC
TACATACTCAATCTACGTAACAACTGGATATATCCTGTAGTTGTTGCCCATTTTTCTGCT
AAATGTTATCTTTAGCACTAAGCATGAGTATGAGGAAACAGTATCTGTGCTCAGATTCCA
GAAATGAAGAAAATGTACTGGAGGTCTTTTGGATAATGGCTACAAGGTCACAGGGACTGA
CTCCTTTGGAAGCTCAGCGATAACCATTTTCAGAGAGAATATGTCAACATCTTTCAGTCT
AGAACTTGATGTTCTGCTGAGATCTAATCTGGGGGTGTCCTACTATTGAATAGGTATAAA
CTAAATAAAAATAGTGAGAGAACATTCATGTGTTCACTCATTCATTCCTTCATCAAACAA
ATATTGAAAGTCTATTAATTGGCAAGCACTCTTCTGACATTAGAAGGAGCAAAGATAAAA
AAGATATTATCATTAACCTCAAGGACATGACAGCATCATGGGAAGGCCAGAAATGCAATA
TGTTAAAGTAAAACACAGTGTAGTGTTTACTACTAAAGAGATATAAACAGAGTACTGTGG
TCTAAAATCATATATATAACATTTGCTTAATGGATGAGAAGGAAACTTTAACTTCAGGAG
GCAGAGCATTAAGAAAGTGAATGACAGGAGGGTCAAAAGAAAAAGCCGACAGTGTTGCAG
AGGCAGGGCATAAAGGAGCTAAACCTTTGCTACCTTCAGTTTTTATTATCCACAGAACGA

FIG. 1M

CAAAGAAACAACAACAACAACAAAACTTTGGATTTGAGGGTTTTTTGTTTTTCTTTTTTT
TTTTTTTCCTCTCATTCCAAGCATCAAACCTTGGGATTTATTTACCTTCTAGCAAACCAA
AATTTATGGGGGCATTCCTATGGTCCTCACCTCACCCCATTTTTCTGTTTTACCTATGAA
ACTTGATCAAAATACTGTCTCCACATTTCTCATAAAATACATTAGTTTAATTTTCTACTA
TTACTTTCTTTTAGTTGATTTAAAAAAAGGTCATTTATGACCTATTTAGGTTAGCATCAT
TAATTTTATCAATGTAAGAATATGGTAGTACAGTGTGAATTCCATTAATGGATATGTTGA
TACCATGGGTTTCTCTGACCTTTCCTCTTCCGCTCCTCCCTGATGATTGGTTCTGAGCTT
ATTATCATGTCAGCAATGAAACAGAAAAGGGAGAAAAATCTCAAGTAGGTTGTCTGTCTC
TTTAACACTGAATAAAGATTTTTTTTTCTCTAACAGACTTAAAAATAGTGCCCTAAAAAT
GTTTTGTTTCATTTGTCTGAATTCCCATTCTTTCCCGTGATCATAGATAGTTGAGCTAAA
AAAAGAAAAAACAAAAAACAAAAATAAACATTGTGTCCTACATTTGTATTAACTTTCTTA
GGAATGAGAAGTAGAATCTTAAAAACCTTAGAATGGGAGTTTCCAAGCTAGCTTGCAGGC
TTGAGTTTTATTGATAATACCTTTAGGATGCATGTATTATTAGAAACATCAGTTATTTAC
AAGTTCACCTATTTAAAAGTCTAATAGGAAAAAATATTTCATGTTGCTAAGTATGTGACT
TCCCTTTAAAAGATAATAATGCTTTCCCTTTAAACAACAATAGTAAAAGAAGTAGAGTTC
CTTTTAAACACATACTTTTATATTATAACCCATTCTGTTTAAAAAATAGCAGGCATATAA
TCTAGAAATGCAAATAATTTAGTGAAATTTTTAAAATTATTCTACATATAATTAAATATG
GATATTCGTTTTCAAATATCAAATAATAAAATATGTCTGAGATGCTGACTAATCCTTAAT
TATAGGTGTGATTTCTACTTCACCATCAATACTATGGTACTCCAAATCTTAACATGAGTC
TGATTTTCTAATAAACATGATGAAAAAAGTTATGGAAAAATTTTGAGATTTACTTTGGGA
GGTTCTATTGTGTTCTGTTCAGCTTCATAATATTCAGTTTCTATGAGTTTGGTATTTAAT
TATGTGTGTTTGTCATTCAGTAGGCTGGAAGTATGACCATTGGGAGATCAAAACGATAAG
ACATTAATGACAGTGCTTTATCACTGAATCTAGTACTTTTTTTAATGAAAGAGATGTTGG
CCTCTTGTATTGTTATAAAACAACACAATTTTATGGCTTTAAATTAAAGTACAATCATAA
CAGAAGACAAAATTAGATTAAAAAACAAACATGGAGTGACTCATATAAAATATTTAGAAA
CCAATAATACAGATAGAGACACATTAGTTCCTCTAGACATTGTGTTTTCCAGTAAAATGA
TCACCAAACTTACCAGGAAAATGATAATTATCAGATTATTTACTTTCAGAATTAAAGGCA
GGAAGAGAAAAAAATGAATGAAGAGGAAACACAGTAACCATATAGGACAATAAGAGTGAA
TGAAGATAAAATGAAAAATCAATAAGATATCGACTTTCTTAAAAGACAAATATCACAATA
GGAAACACCTCAGAAAGGGAAATCTCAAGAAAATAATAAACTGAAAGAAGAAAACATATC
AAAACAACTTGAGGACTGACAAAGTTTTAAAATGTATTTAGATAAAGATACCATGAGGAA
AGTGATCAAGGTGTTCTAGGTAATCACTGAAGATAAAACTAAAAATAGCTTAAATTAAAA
TCAGATAGAGAGAAGGTAACTGAAACAGGCATAGAAAGAAAGTAAGAAGGAATACAATCC
TGAACATCTTAACAATGTCTCAAATGTCAGGAATTGATCCAGTTTTTGGCTGCACAACAG
AGTGGCTATAGTTAACAATAATTCACTGTATATTTCAAAATAACTCAAAGAGTAGAATCG
GAATGTTGCTAACACAAAGAAATGATAAATTCTTGAGGAAATGGATATCCCAATTACCCT
GATTTGATCTTTACACATTGTATGCTTATATAAAAACAGTATTCATGGCCGGGCGTGGTG
GCTCACACCTGTAATCCCTGCACTTTGGGAGGTCGAGGTGGGCGGATCACAAGATCAGGA
GATTGAGACCATCCTGTGAATGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGC
CGGGTGTGGTGGTGGGCGCCTGTAGTCCCAGCTACTGGGGAGGCTGAGGTGGGAGAATGG
CATGAACCCAGGAGGCAGAGCTTGCTTGCAGTGAGCTGTGATTGCACCACTGCACTCCAG
CCTGGGCGACAGAGCGAGACTTCGTCTCAATAAAACAACAACAACAACAACAACAAAAAC
AAAAACAGTATTCATAATAATTAAAATAAATTATTTTTAAAATAAAATAAAATATCAGTA
ATTTAAATTTTTCCTATAGCATAGAGATCTGTAATTAATACTTGTCGATCATTGTTGTTT
CTGTCTTCCCAACAACTACACTCCTGTTTCTTCACATTCCCCCTTCTTCTAACAGCACTA
CATCTTTCTTTAGGAAACTATCCTTTTGCCATTTCATGTATATGGTGGGGTGGGGGAGTT
ATCAATCACAGTACCCCAGCAGATGGGACCAGAGGCAAAAATGCCTGACCTTCTCCCATC
CCCCAACCACAGCAGCAAATGAATTATAATTTGATGCACAAGGAAGTATCGGAGCTTTTG
TGTTGGGTTTTACATATCACCTGTGGGAGATAAATGAACTTTTCCCCACCTAACCTTTAG
CCACTTGGGATGATTAGACATAGAGGTGCCTAAGATCTTTCCCTTTGCCACATTAAAAAC
AAATCATCTATGGCACGAGCATACAAGACCAGCTTTCAGAGACACAAAATGATGGAGAGA
ACCATGATACTAGTTTTAGACCTAGTCACTGAGACTTTCTCTGCTCCTTCCCAGTTACCT
GAGCTTTATTTTGTTTACATTTATCAGATTTGAATGGCTGTACTTCAAAGTACTGATTAA
AATAGGAACCAACCTATATGATTCAGGTGGTGAGAAGGAAGAAAAAGAGAGAAAATGAGG
TTAACAAAAGAGAATAAAGAAAAAGAAAGAAGGAAAACAAGAAACTCTGACTACCTCTCC
TCTTTGACATAGTTTACACTTCTGACAGATTGTTCTTCTCTAAATTTATGTAGAGATTAG
AGTGAGGATGATGTATGCACTGTAGCATGGGTGGTCTTCCAGGAAGCCTTGACTGAATGA

FIG. 1N

GGCAAGGAGTATGTTGCTCCCTCAGTAACCTCAAATTTACCTGCAAGCCTGATAAAAATC
TAACACTAACACTAAACCCAATCTTATCTACAGCCCTAACTGCACCCTAATATTAACAAC
CCTACCTCTGTACTTCAAAACTAAAACTAATTCTGATTTTACTCCCATCTGCCCCTTTTA
CCCTAAAACCAACTGTAAAACTAAATTTAACTCTAAACGTAATCCTAAAACTAAGAATTA
ACTAACAATTTTATCTCTATACCCAACTGTTAACCCCAAGCCTAACTCTAATCCTATCTC
TAACCTAACATTAACCACAAACCTACTTCTAACTCTAATCCTAACCATAACCTCAAATCT
AACTCTGATTCCAATTGTAATCTAAACACCAACCCACCCCTACCCTTTTATCCCAAATCC
ATCTGAAACCCTCATCAGAACACAAATTCCAATTCTACTTCCCACCCTGACTCTGACTCT
AAACATAGGCCCAAATATAACTCTAACTCGAAGTCAAAAACTTAACAAACCTTATCTTGA
AACTCAACCTTTACACTAACCCCAATTCTATCTGTAATCCTAACCCTAATATTATCATCA
AACCTATGTCTAACACGACCTCCAACCCAAAACCAAAACTAACCTCAGACCTAACTCTAC
ATCTAATTATAACCCAAACCCCAGGCTGCTACTTACCATAACCCTGAAACTAAGCTTGAT
CCTTTCTCTTTTTTTTGAGATGGAGTCTCGCTCTGTCTCCCAGGTTGAAGTGCAGTGGCG
TGATCTCGGCTCACTGCAAGCTCTGCCTCTCAGGTTCATGCCATTCTCCTGCCTCAGCCT
CCCGAGTAGCTGGGACTACAGGTGCCCGCCACCATGCCTGGCTAATTTGTTGTATTTTTG
GCAGAGATGGGGTTTCACCCTGTTAGCAAGGATGGTCTCAATCCCCTGACCTTGTGATCT
GCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCACGCCCAGCCGAC
CCTTTCTCTTAACCTACACTAACACTAACTGTAAACCTAGCTGTAACTCTAATTGTAAAC
CTAACCTGATTACTCACTACAAAGGTCCCTCTAATTCTAACAAGAAACTCAATCCTATCT
CAATTCCACCCAACCCCAAAAGTAAATCTAAACTTAAACATAACTCAAAATGTATCTCAA
ACCTTAACCTTCACGAAACTACATGTAGCACTAATGTAACCCTAAGCCCAATCCTATCAG
TAACACTAATGCTAAAACAAACCCCAATCTGTATCTCTACCCCATCACTGACACTACCCA
AATCCCAATATTTAATGCTAATCATTAAGTCTCAAACTAACTCCAAACTTATCTTTAACT
ATATCTCCAACCCTAACCCTAACATTAACCCCAAACTTATCATTAATCATACATCTAGCT
CTAAACCTAACCCCAACTTTAACTCTTACCCTAGCTCTAAAATTAACCCCAACACTATCT
CAAACTGTAATCCTAATGCTAACGTTCAGGCTACTTTTAACCCTCACCCTACACAAAATC
CTGCACCTAAACTCAACCCTAACTTTAAACCTACCTCTAATCCAAACACTAAATTTAAAT
CTGGATCACGACTTGGGCATACTAGCACCCACTAGTGTTCTGGGTGTCATTTCTTTGCTT
CACTCTCATCCAGCTTTCTTTACTAATTTTGGATAATGAATCAGAAAATAGTGGTGTGGA
ATCAGGGTCTTTGAATTATTGTATTATCCAGAGTTTGTCCTGCTGCAAATGATAAAACAC
TGAAAATTAGCTTAACAAGGAAAAAAATAAAAATGTGTGTGCAGGGGAGAAGTAGGAGAG
TTATTGGCTCAAGGAACAGAAGAATTCAGTACTGAGTTTCACAAATACCTGGATTCCCTA
GTCCCCATACTGCCATCAGGATCCAATCTGTCACTCACTCCAATCTCTTCCTCCTCCCTG
TTGGATTCATTGTTAGGTTTCCTGTGGCAAGATGAAATGGCCTCAGGCCTGTAACACAAT
AGGATCAATTACAACAGAAGATAGTATTTCTGTTTTCCTGGTTGCTCAAGCCTAAATTCC
AAGATTAGTTTATATCAAACCTAGTTAGTTTTGCTCATGTAAGGGATTACTGCAACTGGG
TACACTAATATGAAGAGTGGGAGAGTTGGTTAAGGGGGTTCTCTGAAAGGAGAATTAGGT
TACTGTTAATGGGAGAATGAGAAATGGGTATTGTAATGACAAAAACACACACGACTACCA
CAAATGTTGAGGAAGAATTTTCCTTTATGATCATCTAGCCCAACTTTTTAATTTCTAATT
TTGTGGTTTTGACCAGTTTTTTGTTTTTTTTTTTAATGCAGCATGTCATAAAGTTGGGAA
TACTTCACATTTTGTCTTTGAAAATTTGGAGAGTACTTAAAAAGATTTACAAAGGGGAGG
ATTGAATTATTTTAGGAATGTAAATGGTGGTCTTCTGTCTCAGGCAATGTAGATGCTTGC
TAGAAAACAGCTGACTCATGACTGTTTTCTTTCTAATTCATTAATATGAATTATTTCAAA
CTGCAAAGTTATCTCCTTTCTTCTCCTAATCTATCCACTTAGAGTATACATGTTCAAATT
AATGTATTGAACTAATTTTTCTAGTAATACATTCTATGCATTACAAAAATAGCAGTGGGA
AGGTGAAAACAAAATGCAGTTATGCATTTATCTCTAAATGTGTTCAACATCTCTTATGCG
TACTTCAAAATAATTACATTTGTTTAATTTTGAAAAAAATATTAACAAGAAGTTGTAATT
TGGGGAAAATTTAAAGCTGGCGAAAAAGGCTTCATCATAATTGACAATATGGGAAAATAC
TGTATTAAAATCCTAGGTTTCTCCCTTGTTTGCATGAAGGAAATGAAAAATATATAAGGG
AAGGATTTAATCAGTCAGGCAAAAATCTAAATTCATCACAGGTTTATTCACTGCATACTA
TCAATGTGCCCAGTACCTGAATGAATATATTAAAGAAATCCACCTCTTGTACAATGAATG
TAAATGAGCAGAGTGTGGTGATTAAAGGTTGGTATTTGGTGCTGGGTAGACCCAGCTTTG
CCACTTACTGCCCAAGTAAATATTGCCATCCATCAGATATCTCCACCTATCAGACCCACC
CTGTTGTAATAACAAGATTAAAATCTGTATCACTAAAACTTTAAAAGAATTTATAGCCGA
ATCTAGAAATCTTTCACTATAATTTATCTTTCCTTAAAATGTCGTTTTTTTTTCAATTTT
ACTATATATGTTTTTGTACTTGTTTGCTGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTG
TGTGTGTGTTTTTGAAACGGAGTCTTGCTCTGTCGCCTAGGCTGGAGTGCAATGGTGGGA

FIG. 1O

TCTTGGCTCACTGCAACTCCCGCCTCCTGGGTTCACGTGATTCTCCTGCCTCAGCCTCCT
GAGCAACTGGGAGTACAGGCGCACACCACCACACCCAGCTAATTTTTTGTATTTTTAGTA
GAGATGGGGTTTCACTATGTTGGCCAGACTGGTGTTTTTGAAAAGACTTTTTCCTGATT
CAGAAGGTGGGACTCACAATTGTAATTCTGCTAATGGTTGTCTTTCAGTCTATCAATTGC
TTCATAAATGCATCCACTGTTCCTTCTTCTTCTGCCCTGCTTATAATTTTCCATGAGTCC
ATATATCTTTTTACACTGTCTTTAGTCTTATTCACTAAATTAAAAACTAATTTTTGATAT
TTGGTATTCATGACAAGACAATTAGTAGAATTTTGATGCTTCTTGTCTGCAATTACAGAA
TCAATATATTTTCTATATTATTGTATATTCTCTAAATCTTATTTGTATAATAGCTTTCA
GCATGTTCTTTAATTCTGTTTAGATATTTAGAAAGTATTTGTTGTTATTCTGTAATTTAT
TTCAATATTCAATTATAGTTTAATATTTTGTTATCTAGTGTTGTCTTGATTTTGATATAC
GTACTGATTTTGTAGATCCAAATTCCTCTTTCCTATCAGAGAATGCAATTTTTTACTTGG
ATAAATAAGAATCATATCTCCTCTGCTTGCTACCGTATTGCATACATTCATGGGTAGAGA
AAGAGTTAAGCTGATGAGAGTAGGAATTAAGGTAGACCTGTTTGGTAGGTTCTCCCAGAT
TTCAGAGGACAGACATCTTTTTTTCCCTGCCTTGGTCATTTAAACTTTTTGGATTTTGGA
TTAAGTGTAGGCAGGGAAAATGTATCAGATATTTTTATTTTTCTTTGGTGCCATTTGTCC
TTCTCTGCTTTAGGCAGAGAAGCATATGTAGTCCAAGAATGTGCTTTTCTATCCAGCTAC
ATCAATAATAACAATTAGTAAAATTCTACTTAAACTTAGACCTTTGCTGTTCTCTTTTCT
CTGCTTGTGTTAAGTCATGCTCATGATTCTGGCAGTTTTCCACAGTACCATGTACAGAAA
GCTTGAATAAGGTACATCTAGAATACTCATATATGTTCACTTCAAAAACACATTTTTGTG
GAATTCTAAATGCAAATCTCAATAGTGCAATTCTAATTTACAATGAGAAAAAACTAAGGG
ATTTTTTCTGGTGATTCTTTTTGCTCATTTATAAATATGTTTTTAAATGGTAAGCAAATA
TATAAATTAAGCTTTTCCTTACGTAGCTACATTGATTTACTAGTGGTGGAAAAGGTTAAG
CAAAACTAATTTTCATGAGTGTAAATGAATTAGTAAGTGACATATGCAATGCTTAAGGGG
AATTTGCATAAATCTATGACTGATACTCAACCTCTTGCTTAGCGAGAAGATAATTAAAAT
ATTTTATACTTCAAGAAGACCTAGTTTTCCAAATTATTTACATCCACAAACTCAGATTTT
ATAGCAAGTAAGAAAAGTTAAGTCAGAAGCATATACTATTAACAGCTACTTACATTGCTC
AAATTTAATATACGATTGCTGCTTTTGTTGGTTTTGAAATGTTTCTTGACCATGGATCTG
AATAATGAAGTTATTCAAGAAGCAACTTTAAGAATGTTATATTCTTAGAAAGAAGCTATA
GATACAATAATATTAAAAAATTAAATGTAAGTTCCTGCACTCACAGTAGAGGTAAGTTCA
AGGTTATAAGAGAGCTTATAGATTCTGAGATTTGGAAAGAAGAGAATAGAAAAAACTTTT
CAGATTAAATAATGTGTTAATTGTGCTTCTAAAACAGCTTTGGTGATCTTAATAAAATAA
ATATTGTTTTTATTTCCATTTTTGCTTTTCAGACAAGAAATGCTACTTGATGGCTGCATA
TATTTGTTTTGTCTCTTTTCACCACCTACTCTTGCTAAATACTCTCAACCCACTCATGAA
ATTAAAGCACATTGGAAAACATTTATCAACTACCTGTAAATACAACCTATGCTCTCTTTT
GTGGAGGTGATAGACATTCATCAATGGAATAGTTGATCTAAATCCTAGTCTTCATTATCT
TGTTTTATACATTCTTGTCTTAATCAGTTTGGGCTGCTCTAACACAATACCATAGACTAG
GTGGCTGATGAACAACAGAAATTTGTTTCCGACTGTTTTGGAGACTGGGAAGTCCAAGAT
CGAATTTTATGTCTGGTGAGGGCCTGTTTCCTAATTAATAAACATCTGTTGTCTCATATG
TCCTCACATGATAGAAGGGGCAAAGGAGCTCTCTGATGTCTCTTTTTTAGAATATTAATC
TCGTTCATGAAGGCTCTGCTCTCATGACCTATTCCTTCCCAAAGGGCCCACTTCCAAAGA
CCATCATATTAGGGATTAGGTTTCAACAAATGAAGCCAGGGGGAGGTTGGTAAACATTCA
ATCTATAGCAATGCCTATCTCCAGGAGCTGCCTGTGGAAACACTTTTATCTGATATGGTA
GTTTAAAGCATGGCAGGGATAAGTGGTATGAGGAAAACTCTCCCTGCCACCCAACGCACA
CATCCCACTTAAGCTTCAGCAGCTCCAATTTTATCTGTGTAATATTTGGTTCCACATCAA
AGTTGTTTTGAATATACTTCCATTACCTTAAAAAATGTAAAAACACTGCTTTAAAAAGCC
AAGCCTATTCCCTTTTCATTATTCAGAGTTCTTCCAGTTTTACCGTTACATCAAATTAGA
ACTACATAATTAGGAACCCCTCTCTAAATTTGCCTCTATACAGAGAAAAACTGTGCCTGA
AACTTTATTAAAACTCAATAAAGGAAATATGTATGAATGTATATATATAATTTCTCTGAA
GGACAGAATTTGTACTTCGTTCCATACATAAAAACTCATTTGACAAATAACAAGCATAGC
TCCAAGCTCAAAGAATAGCTTAATTTTTCCTGATTAGTTTATATCTCTCTTATTAATCAA
TGACATTTAATATTACAACCATAGCTTGGGGTTTTAGTTTATTTGCTTTCTATCTTTTTT
ATACTGTCGGCCTACCTGTGCCCAACTATGTTATAGTCAGGGGTTGGTAAAATAAAGACA
AAACAAATCCTGTCTTCCTGGAGATCACCTTCACTGGGGGTTGAGAAACAATAAGAACAA
GTAGTAAGTAAAATATGTACATTAAAATTTTAGATGAAGTTAAGTGCTATGGAAAAAAGT
AAAATGGAAGAGGTGTTATGGAGTACCTGTTCGGGTATGGGTTCAATTTACAAGTGGATG
GTCACCTTCTCACTGATAAGGTGACATTTGAGCAAAAGTCTTCAGCAGGAAGGGAGAATG
CCATGCAGTTATCCTAGGAAAGAACATTTCCAATATAAGTAACAGCCAGTGCAAAAGCCC

FIG. 1P

TGATGTAGATGCATACCTTAGGTATACGAGTAACAGTAAGAAATTAGTGGCACGAAAGAC
AGATGTACTTGGAAACCAAAAAGAATCTCTGGTAAGAAATTGTAAGTCATTGTAAGGACT
TAAGGTTTTTTTTTTTCCTCTCCAAATGAGATGGAGATCCATTAGAAGGGTTTGCGTAGA
GAAATAATATGATCTGACTTATATTTAACAGGACTACTCTTTTGCTGAATTGAAAATTGT
CTCTAAGGGTGTATATCAGATCTTATATTGATCTTACCCTTCTCTGTTCAATATTTAACA
CACAAGCCTGTTAAATAGTCCATTCCCAACTTCTGTGACTTCTTGCTTGAGAGCCTTTCT
ATCCCCTCTCATAAGGGCTGTGAGGGCCTAATCTGCTTACCTATCCAGCAGGCTGGGAAT
GACACAGAGCACTCACCAGGAGCACTCTCAACCTATGACTCATGGAAGTTGGTAGATGAA
TACCCCAGCTCTCATATTCCTTGGGTGGAAGAGCTCTGAGATGTGTGTTCTACACCATTA
CCCAGAGGGCACCCTCTGGATTAGGCTCAAGTTGCTGACAGTAGTATCTTGCTGACTAAC
ATAATTTTTATTAATTTTCTCCCCATTTGACCTTATTTCTCCATTTTTCTAATAGTGTTC
ATTGGTATCACTTCCAAAATAAATTACCTTTACTTGAATATTTTTCTTAGAATCTTCTAT
ACAAACCTGAGCTAATACTGGGGCAAAGAGTGGAAGCAGGGAAATATTTTGTAGGTGTTG
TGGTGATGTAGGACAGAGCCTGATAGCTTGGATCAAGGTGGTAGCAAAGGAGATTGTAGA
AGCTATCACACTCTTTATATATTTTGAAGACACAGCCAAGAGGTTTGGTGGAAAAAATGGA
TTGTGAGAAGTAATAAAAAGAGTGGGAGAGAAAGTCAAGGATGTCACCAAAGTTGTCCTA
AGCAAGTGGAAACTTAGATTTGGGAGAATCAAAAATCCTAAAATATCCAAATCCTCTCCC
CTGCCTTCCCCTCCCCTCCCCTCCCCTTCCCTTTGGAGATAGGGTCTTGCTCTGTTTCAC
AGGCTGTAGTCTAGTTTCGCGATCTCGACTCACTGCAGCTTCGACCCCCTGGGCTGAAGT
AATTCTTCTACTTTAGCCTCCCAGGCAGCTGGGACTACAGGATTGCACTAATGTGCCCAG
CTGATTTTTTTAGTTTTTTTTATTTTTAGTGGAGATGAGGTCTCGCTATGTTGCCTGAG
CTCAAGCAATCCACCCTCCTCAGACTCCCAAAGTTCTGGGATTACAGGTGTGAAACACTG
TGCCTGGCCCAACATTTTATTTTCAAATATTTAAGTTTTGAATGTCTATTCGATAACCAA
GTAAAGAAGTCAACTAGAATATATGAGAATGGAGTTTTCTAGAGAAGTCTGGGTTGAGGA
TGTACTTTTGGGAAATGGAGCACATACTTGGTATCTAAAGCTGTGAGCCGAGATGAGATC
ACTAGGTAGGTAAATATAGATAAATTAGAGAAAATATCTAATAATTGAGACATGGAGTAC
TATCATAAATTTTGAAAAGACAAGAAAATGTGAGAGATCGAGAAGAATGGCTGGGGAAGA
AGGAATCTAAGGTAGTGAAGAGATTGAAATGTGTCAAGGAGAGAAGAGAGTAATTAGCTC
AAATGCTACTGATAAGTAAAGTGAAATGTAGAATGAAAGTCAACCATAAAATTTGGCATT
ATGGGGATCATTAATGACCTTAAAGAAAGTGCTTTTAGTGTAGTAATAGAAAGATGCAGA
AAGTAAGTAGAGTGAATTCAAATTCAACAGAGAATAGACAGAGAGGAATTGAAGACATTT
ATACTGACAATTCCTTCCAAGACTTCTGCTATTAAAAAAAATAAAAAAAGAAGGAGAAAT
GGCAAGTGTTTGGAGGCCAATTTATACTCAAGAATAATTTCTTGAGTTGGTTTTTTGTGT
TTGTTTGTTTTTGATTGGTTAGTGTGTTTATTTTTTAGACGGGATTGGAGAAATACTTTC
ATTTGTGTTTTTACCCATGTTTTCAGCCTTGCCCTGGCTGCCTGGTATAACGCAACTCTA
TTTGTTATTCTGCTATTATAGTTTCCCTAGCTTGAATTTTTTTACACCCTTATTATAATT
GTAGCGTTGCATGCCTATTTCAAAACATCTCATGTACCCCATAAATATATACATCTACTA
TGTACCCACAAAAATTAGAAATAAAAAAATTTAAAAATTATGATTTTTTAAAATTTGTTA
AATAATGTTTACTGACTCTTTTATTTGTTGAAATCATTCATTTTTTGGAATATCAGGTCC
AATTAAATATTTAATCAGACTTTGAGAAGGATTTAATAAGACCAATAAATAACCAAGTAT
TAGTTGAAGGAAATTTCAGATATTTTGGTAGCAGAAGGAACTGAGTTATGGCTCAAGAGT
TTTTTAATAAGTGTGAGTGGAGTTATACAAACTACTCATTAAAATCTTTATTTGAATTTG
TAATATCTGAAACCATTTTCATATTGAAGAATCACTTAAAATAGTCATAAAATGTAAAAT
TGCAAGACAATTAAAAACAAAAATATGATTTCACGACTGTGATAGTACCTGAGAAATTTC
TTCATCTCCTTAGTAAGAGAAGTATTACACCTATTTATAGTTATTTTATGAAACTAGCTA
AGATGAATTATGTAGAAAAGATACAGATTTTCAAACAGAAACTAGAATTAATGGAAGCTA
TGTGAGACTATAAAGAGTTTAATAGTTATTTTGATTTTTTTTTATGAGTGCAAGGAGTAT
AGCGAAAAATAGCATCTACCTATAAGGATTTGCAAAGCCAGTAATCTTTCTAAAAATATC
AGCAAACCCAGAATTAAGGCTTATGTTCTTAGCTCATTGTAACTAGATCAAAAATAAAGA
AGGCCAAATAAAGGTATGTGACATTTGTTGAAAACCTGAAGTGTCCTATATGCAGAAATA
TTTTTATCATTTAATTAATTTCAGAAACTTCTTAACATGACATGATCCTCTTGAAAAGAT
CACATCAAAAAAGGCAAAATAATTGCATAATTATTGTAGAATAATTTTTGTGTGAGTATT
TTTGACTTAGTGTAAGTTTCCAAGTTCAGGATTTATCATGCAGTGAAAAAAAATACACTT
GTCTAGAAGACAGGAGACTTCATTATATTCCTCTCTTTACAATTAATTAACGTAAGACCA
TTTAAAATATGCCTAATTTTCCAGGCATTGGTTTGCTTTGCTATAAAATGGGAGGATAGA
AAATAACTTTCAAAATATCTTATAAATCTAAGAATCTTTGCATCTTATAAATCTAAGAAT
CTTTGGAAATTCATAGATTATTGAGATGGAGTCTCGTTGCTATGCATTGTAGCAAAGTTG

FIG. 1Q

GAAATAAATTCTAAATTTTATTTCATTTATATTGATCAATAAATTGTTACATTTCACTAA
TACAATAAGGAAAATTTATTTTACCTGAGTGTATGTCTAGCTTGTGAAATAAAAATGCTC
AATTATGAAAGCATTTATTGCCATTTTGAATGAAAAATGTAATATGTAGAACAGAATTTT
TTTTGCCTTGAACTCAGTTAAATGTAGAAATTGATAAGGACTTGCATTTTCATGAACTTA
ATAATTATCTGTCTTTTCAATGGTCTCCATATCAAGTCTGAGAAATATGGATGTGATTTA
TTTTAAACCTCACCATTTGAAGTAAATCTAAAGATTCCATTAGGTTATGAGCATATAGGA
TACAAGGACCATATTGACAGTTTTGTGGGATTGTATTAGGATAAAAGGGTAGGAACAATG
GGGAGAAAATTATAGCTTACAATAGGGAAGAACCAAAAATTGTTGCAAAATGATGGAACA
GGCTGAAAGAATGATATAACCTCCTAAACACTTCAAATGTTTAAGCAGTTCATTGTACCA
GGGCCATTGTAGCAAATATTTTCTGTCTTGGGTGGAAGGTCAGTCAAGGTGACTGATAAA
GTTTCTTCTAACGATAAAATAGCACAACTCACTTTTTTTCTAACCTCTAAGAGTATATTA
ATATCAAAAGAAGGCAAGCAACAAACTACTTCTGAATGTTAATATATATCTGCATTCATT
TTAAAAGTCTGCTACAACTACAGATAGAGGAACAGTTTGTAGTATCCGTGATCCTAGAAC
AAATTTAGCTTTTAATATCTTGTCAACTTTTTTGTTTTAGTATCTCTTCCTTGGAACTAG
CTGAGCTTTAATGGCATCATCATGTGATATGACTTGAGATTTATATTTGGAAGAGCTTTG
AAAAATCACGGATTGTTACCCTAATGAGGTGTTATTCAGTCTTTTAAACAAGAGCAATTT
CTTTACAAAAAGGAGCAGAATTCTTAATTGTATCTGTAAACCTCCATTTAAGAATGAATT
ACTTGGCTGGGCATGGTGGCTCACACCTGTAATCCCAGCACTTCGGGAGGCAGAGGCTGG
TGGATCACTTGAGGTCAGGAGTTTCAGACCAGCCTGGCCCAACACGGTGAAAAACAGTCT
CTACGAAAAATAAAAAAAAAAAAAAAAAAAAAAAAATAGCCAGGTGTGGTGGTGTGTGCCT
GTAATGCCAGCTACTCGGGAGGCTGAGGTGAGAGAATCACTTGAACCTGGGAGGTGGAGG
TTGCAGTGAGCCAAGATTACACCATTGCACTCCAGTCTGGGTGACAGAGCGAGACTCCAC
CTCAAAAAATAAAAAATAAAAAAAAAAAGAATGAATTGCTCATAAATGTGCCTCACTGAT
GATTAAATTTAATCCTGCAAGATTATGTCTTTTGATGGAAATGAGAGGGTTTATACAAAG
TTTTATTCGTGATGTTATCTATGTCATCTATTGATTTCTGCTCTGATTCATGTGGATGAA
GTTACACCTCACACTTTAAGCTGGTGTCAGTCTTCCCATTTTCTGCTGTGATGTGTACTC
AAGATCTCCAGATTACATCTGTAATGTAATGCAGCCATGATTGTTTATAGGTACATTTAG
ATGAATTCAATGATGAGTTATGTTGTAATAAGTGTCAGATTTAGATGAACCATACAAATA
AAAGAACCATGCATTAAAATGACAAATGTGTAAAAGCATTATTTGGGCCTTAAGTCAAGG
CCCAAATGTGGATACTGGTACTGAGACATCTTTCAGAAAGGAGGTATGAAGTACTGAAAA
ATATTTACAAAATGAAGACTACTTTTATCTTACTTATCATGATTCTTTTATTACATATGC
ATTTTCTAAGATAACTATAGTGCATTAGTTTGTACTATGTTAATATAATAATAGGGTAAA
TCAAACAATGTTTTCTAAATCCATTAAAATAGAGTTCCCTAAGGGAGTTAAAACAATTAC
GTTCTACTGTATATTATTGGCATGCTTCAGGAGACATGATTTAATCTCTAGACTATCAGA
ATTCAAGAACTAGTGAGTCATATAACAAAGGAGGCTTAATCATGCCATTTAAGTGTCATG
GAAAAAGGTTTATTGGTCAGGAAAAATTAATTAGAAAAAAGTTATAAAATACTTCACTAA
GAAAATAAAATGTCAGGAAGCCCACTTAGACAATGAGTGAAAATGAAACAAATTCAAGTT
TTTACAATATTTGGTTTCTATAGGATTGCTTCATTGTTTTGGTTTTTGTTTTTCCCCATA
AGCTGATCTCAGAAACTTTTCCTCTACATGAAGAGGCTGTCATTTTTTCATGGTGTGTGT
TTGTTCACATGCCACACAGACAATCAATTATGAAGAAAGGAGAGACTCGTAGGAGGCAGG
GCCAGGCTGTTCACACTTTTAAACTAGGTAGCCACAAATGAGGCTTAGTTACAAAAACTT
GAAAACTGGATTCTTCCCAATGTATTATACATCCCCAAAGAAATGATGAAGTTCCTTACT
CTCTTCTCTTTGTTTTTGTAAATCTTACCACTTCAAGTGTTGGCAATACTTACTTTAAAG
TAGGTTTTCATATTGGCTTAGATTTTTTTTTCATTAACTTGCAATTTGTGGTTGGGAAAT
GATCTGCTTTTTGTTTCAGGTTGTTTAATGTTTTCCAATGTAATATTCTTCTTGCACTCC
AGTGAGTTTATTTACAAAACATTTAATGTCATTTGCGTCTTCGAAGAACAATGTATTCGG
TTAGAACAAAAGTGAGCTCCTGCATAGAGCTTATGATGGTTTATAATTGGTAAATTATTA
CCTTGGTCAAGTTTGTAAACTAATAAAGGGAGTAGAAAACTTTTAGATAAAAAAAACTAC
CTCATTCAAAGGGACCGTTCACCCACAAAATGCCTTTTTGTTTATCTTTTGGAATGACAC
CATTGGAAACTCAGTATGGCCACTTTTATGGTAATAATAAAAGTCATATATAAAAAGGAT
TATTAGAAATGTGTTATTTCTTAGGCAGGTATGCTTATTTAAAGTATGTATGCATACATA
CTTTAAACTACTAAATACAAATAAATTAGTAGTACAGTCATTAGGATTGCTCTTAGTTTG
TTAGTGTTGGAATAGACTTTTGGATTTTCTTCCTAGCTTAGATTGATACAATGTGATGGG
GACTTGCTCTCCAAACACAGGAATAGGTGGCCTGCAGACACACTCTGTGATGCTGTAATT
CTAATCCTCACTGAATATATCAGGGGTGGACATCTGGCCTGGGGCAATTCAGATACTTTT
TCTTAAAATTTATACTACAAATTCAAAAGTGGTAACTCATCTCTGCCATCACTTATAGTA
GAATAAGACCCACTGTTGCAGTGGGGAATTGAGAAACCCAGTCCACAGGGAGAACAAACA

FIG. 1R

TGGAGAATAAAATAAGTAAATTAGAACAGGAAAAATGCCAAAACACACAGACATGACCCT
GATAGTTTTCCATTTCCTGATCACTGTCCCTTCCTGTGGCTGGATAAGGAACTGTCTCTA
GGCTCTGTAAGACATATTTGCATCCTTACGACAAATTTCTACTCCTTTTCATAAACTAGA
CTTGGGTTCTTTAACTTGCAACAGCAACAACAATAAACGATTTTGTTGGGTACAATCTGA
TTTTATTAACTTCTGGATTTAAAAGCCCTTCTAAATGTTGATTGGCATTGTTTTTACTTC
CTAAGAGTACGCTCATGCACCACATAGTGATGTTTTGGTCAACGACAGACTGCATTTACG
ACTGTGGTCCCATAAGATTATAATACCATGCTTTTCTGTACTTTTCTATGTTTAGATATG
TTCAGATACACAAATGCTTATCATTGTGTTATAATTGCCTACAGTGTTCAGTACAGTTAC
ATGCTGTACAGGTTTATAGCCTAGGAGCAATTGGCTATACCCTATAGCCTAGGTGTGTAG
TAGGCTATACCATTAGATTTGTGTAAGCATACCCTATGATGTTTGCACAATGATGAAATC
ACCTAAGGATGCATTTCTCAGCATATATCCCAGTCATTAAGCAAAGACTGACTCTATTAT
TAGGTCTATTTTATTCTATAGCATTTGATCATGAGATATGTGAAAATAAATATAATTTTT
AGAAGTACAATAACTTTCAAATCCTGAATGTTCTGTACTTTCCATCTCACAAGCATTTTG
CAAAGCATCAAATGGTATAAGCCAGATTACTGTTAAGGCAACTTGGAATTAATATGCTGC
TCAGTTCTGGAAAAGGCATATTCTGTAAATATAGATGAGAGAATATAGACTTTTTCCCTC
TCTTCTTACAATCCACATTCTATTCAGTATTTCATTTACTTGAGGGGTTATATGCTACTT
ATCTTTATCTGTTGTGGAGTGAGGACACATTCCAAATGCCTTGGTATTATTAAAAGCCCT
TCATGATGTGGCCCCATCTTTTATGACTTTTCCTTTTCAACTGTGCCCTCTAGCCTTATT
TGATTTCTCTCAAATTCTTAAACACAGCATGCTTCACTGACCTTTAAGCCTTTGCACATA
CAGTGTTGATGTGGAGCTTCCTGACCAACTCCTAATTCTCCTTCAGGCCTCAATTTAAAC
ATCACTTCCTCTGGGAAGCTTTCTATTATTCCCAAGGTACTGGGATATGTTCTTGCACAG
CATGCTGGGCTAATGTCACAATGGCTACCTTGTTTTATTGTTAGTATTTGATCAGCGACA
CCTTGCCAGGGAGCCCCTGAGTATTGTCTGAGCAGAAACTATGGCTATCTTGTCCCCTGT
TTAGCACAGGGCTTCTCTAAAAGTGGGCTTCTCTAAAAGTAAGTGCTCAAGAACAACAAC
AAAAAGTGTTACATTAATAAACACACACACATACATACAAAGAAATACCTGTCTTTCTCC
ATATCTCAAGATCATGCTGAAAAGCCAGCATTCATGAACAAATTCCTGTGCGAAGATTGA
GAATGAAAGATGAATAAGAGGTATCTTTAGAACCCAATTATGGCTGCCGTTGTTCCCTGA
GTGTGAGGCTTGCTGTTAGAGTGACAGAAGGAATTTTGACTACTCAAGACCATACAAATT
TGGAAATGACTCCAAAGTAAACATGGTTAGATAACTACACATTCCATTCCCCCTTTTTTA
TTTCTATAGAATCCCAACTTTGTTCAAGTAGTAACATGCCCAGCTTCAGAAATGAGTCAT
GATTTTTCTAAAGCAACAATATCAATCTTCTTTCCCTTCCCCAGTGATTGGTATGGAAGT
GGACATTTCAGCAAGTTTTAGCCAATAACGTGAATTCTGTTTTGAAGCATCTAAGAAAGA
TTTTGCTTTCTGCTGTAAATCAAAAGCAGAAACAGGAGAAGATTCTTTTGGGCCTCTTTC
CCTCTTCCTGGCGTGGAAGTAGTTGTGAGAGCATATGATACCCAAAGTTTCGGTAGACAT
TTTATAATTATGTGATGAATAACCTAAGGATAATTAAACATATAAAAGAATGGAGAAAGA
CTGAGTCTGTTTTACTCCACAAGATGCTGAACCAACCCTGAGACATAATTTATCTGGATT
CTTAAATAACTAGTGTCTTTGTGGTTTAAGCTGTTCTTTGTAAACAAACATATCATAAGT
GATTAAGTGATGTTATCTTCCTTTAAGGCAATCAAAATGCATCTGACAAATGGCCATCTA
ATTTAAAATTCCAACTATGTAGACATCTCAAACAAAGTCAGTATCTCAAAAAATATACTA
CAAAAATTCTCATGTGTCCATTGGGGATAACTTCCAATGCTCTTTCATTGGTATTGTAGC
TATGGCATTTGATTTCCAATTGTATGTGGATCAGGTAGTTGCAGGGTGACTCTCAAGGGC
GAGAAGAAAGTAAGAGTACATGAAAAAAAAGAGGAAGAGAGAGAGCAGACAAGAAGGAAG
AACAAGACAAAGTCAAACCCTAGGTAGAAATAAGAAGGAGCTAGTACAGAAAGCAAATGC
CTAAGGTGTTGGAGAACATAGAAAGGTAGAGTGGAATGAAAAAGAAAAAAACACTAAATA
GCAGCACATAGAATCTTGGGGTTTCAGGGATATTGTTTATGAAAGGTTAGAATAGGCAAC
AATCTACCTTGTGGCATCTTCTTAAAATTATCAACATATAAAACAAACAATAATTATTTA
AATTACCTGTCATATGGGTCTTGTCATTTATTTATAATTTAAGGAGAATTAAAACTGAAC
TAGTTGCTGGGGAGTGACATCAGCAAGATGGAGATATAGAAATCTTCAGGACCTCCTTCC
GTCCATGGAACCACTGACTCAAAAATGACAAATGGAAAAAATTTACTTTCTGAGAAATCA
AGAAGCCAGTTAAGAGGCTCCTGTATCTCAGATGAGTGCAAAGCCAGCTGCAACAGAGCC
AGCAGAAAATTTGTTGTACTCACTCTTCATGGTCACTTCTGGCATAGCACAGTGCAATCT
AGAAGAAATTCTCGGCTCCTGACTACTTTCTTGGAAAAGAAAGAGAAAAATGTACCATAT
GTCTAATATTCTGATGGGGATGGGGTGTGGGCTGCTCAAAGGACTAGCTTCCGTCATGCC
TAAATACAAGTGCTAATTGGGAAGTCCACAATGTTGGGGGCTGCAGAAAACAAGGGCAAC
AGTTTGGACTAGCATGCACTCATTTGCCGCAGTTCCTCCTCTCACTTCATAGAATGAGTA
GAAGAACCCTTAACTCTCAAGGTTTTTTCCTGGGGAGAGAAAGAGTCAAAGCAATTATA
CAATATTATGGCTTTGTGGGAGTGATGTATCCAAAAAAAAAAAAAAATGAGTTTTTACCAC

FIG. 1S

ACCAATCTCAGAGTGCAGATGGAACCTAGCATATTCTAGATGCCTGGGGGCCATTGAGAA
CAAAAGAGAGCTAGGCAACTTTCAGCAGCTCCAGAAGAACTGTGGTACCACAGATAGACA
CCAAAGGGAGGAAGAGATTACAAGCTCCTGAAAAAAGAAATGAGCAATTCATTCTAATTG
AGAATTTACACACACTGGTACAGATAAGATGAATTTGCAAAAAAAGAATAGAGGCCCCAG
AATTTCTAGCTGGGTTTTTTGGTGAAGGCCTTTCTCTGTATCAAGCTAGTCCCTAAAGAC
TGGGTGAGGTGGTTTTTGTTTGTTTTACATTTTTATTTTAAAAGATGGGGATCTCACTTT
GTCACCCAGACTTGAGTGCAGTGATGCAATCATAACTCACTGCAGCCTCAAACTCCAAGG
GTCAAGTGATCTTTCCACCTCAGCCTCCTGAGTAGCTGAGACTAGAGACACATGCCACTG
TGCTTGATTAATTTTTATTTTTTTATTTTTTTTCGTAGAGATGTGGTCTCACTTTGTTGT
TCAGGCTGGACTTGAACTATTGACTTCAAGGGATCCTCCTGACTCAGCCTCCCAAATCAT
TGGGATTACAGGCATGAGCCACCATGCCTGACCTGTTTTGTTTTGTTTTAAAAAACTCAG
AAAAATTTCAAAATAGCAATTATAAAGACAATGAGCTTAGAAAACCAATTAATGGACAAA
ATGTAACTATAAGTAAAGAGATACATGTAAAAGAATCAAACAAAATTTGCAGTGGAAGA
ATATGATAACCAAATTGAATATTACATTAGAGGAGTTTAATACTAGATTTGAACAAGCAG
AAGAAAGAATCAGGGAACTTGAAGATGGGTCATTTGTAATTATTCAGTCAGAGAAACAAA
AAGAAGACTAAAAAAGAGTGAAGAAACCCTAAGGACATCATCAAGTAGACCAATATGTGT
TATCAGAGTTTTAGAAGAAAAAGACAGAAAAATAGGCATAAAGCATCATTGACAAAATAA
TGACCCAAAACCTCCCAATTATGAAAGACAATAGATATTCTGAATCCAGAGCACAATGGC
CTGCAACTAAGATGAACCCAGAAAAGTCTATACTTCAGCACATTATAATCTAATTATCAA
AAGCCAAGGACAAAGAAGGAATTTTGAAAGCAGAAAGAAAATAGTGACTCATCAGATACA
CAAGGGCTGTCATGAGAATATCAGCAGATTTCTCAGCAGAAAACTTGCAAAACAGAAATA
AGTGGGATTACATATTCAAAGAGCTGAAAAAAAGTCTGCCAACAAAAAAATCCTTTATCCA
GAAGAATTTTCTTCAAAATGAAGGAGAATAAAGGATATTCCAGATAAACAAAAGCCAAGG
GAATCCATCACAATTAAACCTGCCTTACAAGAAATGCTAAATGAAGTTGTTCAAGTTGAA
ATAAAAGAACGCTGAACAGCAACACAAAAGCATATAAAAGTATAAAGCTCATTGGTCAAA
GATAGATATAAAGGAAAAACAACGGGATATTATAATGGTGGTGGGTAACTTACTCTTCAT
CCTGGTATAGAAGTTAAAAAAAACCACAAGTATTAAAATAACTGTAACTATAAAATTATT
AATGAATACACAATGTAAAAATATGTAATTTGTGATACTGATAACATACCATGTGTGGAG
GGGAGAAGTCAAAGTGTAGAGTTTTAAATAAGACTGAGGTTAGGTTTTTATCACCTTAAA
ATAGATTGTTATAATATGTTTGATTTAAGCCCCATGGCAACTACAAAGAAAATACCTACA
GGTAATAAACAAAAGAAAATGAGAAAGAAATGAAAGTGTGTCTCAGTCCATTTTTATTTT
GCTATAACTAAACATCTGAGACTAGGTCATTTATAGAGAAAATAAATTTATTTCCTGCAG
TTCTGGAGGCTGTGAAGTTCAAGACTGAGTTGCTGCCTCTGTTGAGGGGCCTTCTTATTG
CATCATAACATGGCAGAAGGCATCACATGACAAAAAAGCAACAGCAAGAGCCAAACTGGC
TTTTATCATAGGCCTAGTTTGTGACACCTTACATAGTCCTATGAAAACCCATTAAGCCAT
TAGCCCATTAATCCATTAATTCATGAATAGATTAATACATCCATGTGGGGAAAGCCCTCA
TGACTCAAACCTTTCTCAAAAAACCCATCTCTTAATACTGTTACATTAGTATTAAGTTTT
AACATGAGTTTCAGAGTCTAGAAATATTCACACCATAGCCTTTCACCCATGACCTCCCAT
AATTTATGTCCTTATCATATGCAAATACCTTCATTCCATTCCCGTAGCCCCGAAGTCTTA
ACCTGTTCTAGCACCAACTCTAAAATACGAAGTCAAGAGTCTCATCTGAGACTCAAGGCA
TGATCCATCCTTGGGCAGGTTCCCTTTCAGTTGTGAAATCAAAACAAGTCATATAATTCT
AAAATACAGTGCTGGTACAGGAATAAGACAGACATTCCCTTGTCGAAAGGGAAAATAAAC
TAGAAGAAGGGGTTAATGGTCCCCAAGCAAGTCTTTAACACAGCAGGGCACATATTAAAT
TGTAAAGCTAAAGAATACTCTTTTTTGGGTCCATGTTAAGCATTCTCTGCACAATGTGGG
GAACACATTGAGCCACTCTGCCCCTATGGCTTTGCTGTGCTCAGAACACACTTCAGCTTT
CTCAGATTGGAATTGCTCATTGGTGCCTGCAGCTTTCCCAGGTGGGCACTGCACACTGCT
GGTGTTTCTATAATTCTAGGATCTCAAAGGCAGCTCTGGCTCTCACCCCGTATTTTTACT
CAACATTGCTGTAGTGGGGCTCTCAGCCATGGCTCTGTCCCTGTGACAAGTCTCTGCCTG
GGTCCCCATGCTTTTAGATACATCCTCTGAAGTCTAGGTGAAGGCCATAGTGGCCCTACA
ACTCTTGCATTCTGTATCCCTGCAGAATTAGCACCAGGTGGACACTGCCAAGGCTTATGG
CTTTTGCTTTCTGGAGCAGTGAGGTAAGCTACACTTGGAGCCTCTTGAGCCAGTTGGAGT
GGCTGAGGAATGATGCGCTCACATGAAGGGAGCAGAGGAGTCCTGAGCAGCCCTGGGCAG
CAAGCTGTGGAGAGTACCCTGGGCCTGTCCCCTGAAACTATTCTACCCTCCTTGGCCCCT
GGGCTTTTCATGAGAGGGGGCAGTCTTAAAAATATGCAAAATACTTTTCAAACATTCTCC
TCATTGTCTTAATGAATAACATCTGACTCCCTTCTATCAGTGCTAATCTCTTTAGCAAGC
AGTTTTGCTGTTTACATGGCTAAGCAAGCTGCAAACTTTTCAAATCATTTTGCTGTGATT
CCCTTTAATTATACATCTGTCTTTAAGTCATGTTTTTGCTCCTGAATTGGCCAAAAGTAA

FIG. 1T

CCACACAGCCAAAAGTAGCCAAACAGCATCATGAATGCTTTGCTCCTTAAAAATTTCTTC
TATAAGATATTTTACTTTATTATTGTCAAGTCTGGCCTTCTACACAGCCCTAGAGTATGG
ACACAGTTCCAGTAAGCTTTTTGCTACTTTATACCAAGTATGACCTTTATTCCAGGTTCT
GATACCTTGTTCCCCCTTTCTGTCTGAAACCTCATAACGGCCTTCATTGTCTATATGTTT
ACTAGTATTTTGGCCATAATCACTTAAATAATTTATAAAATGATTCAGACTTTCCCTAGT
CTTCTCATCCTCTGATCCTTCACCAGAAGCACCCTTAACACTCTATTTACAGCAATATAA
GATTTTTTTTGCCTGCTCCTCCAAACCCTTCCAGCCTTTGTCCATTACCCATTTCCAAAG
CCACTTGCACATTTTTAGGTTGAGCATCAGCCTCACTTCTTGTTACCAAAGCCTGTATTA
GGGTTCTCCAGAGAGACAAAACCAATGGGATATACAGAAGGGGATTTGTTAGGGAAATTG
GCTCACACAGTTATGGAGACTGAAAAGACCAAGGTCAAGGGGACGTATCTGGTGAGAACC
TTCTCATTGTATCATAACATGGCAGATGGCATCACATGCTAAAAGAGCAAGAACAATAGC
CAAACTGGATTTTATAACAGACCCACTCTTGACGACTATCCTATTCCTGTGATAAGCCAT
TAATCTGTGAATCCATGAGTAAATTAATCTATTCATGAGGGCTCTGCCTCTATTGTCCCT
TAAAGGCCCCACTTCTTAATACTGTTACATTGGGGATGAAGTTTCAATATGGGTTTCAGA
GGAGACAAACATTCAAACCATAGTGATGTCACTACAAAAAAAATTAATGAAACACAAAGGA
GTACAGTAAGAGAGCAAAATACAGATAAAAGTGCTATATGATATATAGAAAACAATAAAA
TGGCAATAGTAGGAGTTTATCTGTCAGTAGTTACTTTAGCCATAAATGAACTAAACTCAA
ACAAAAGACAAAGATTAGCTGACTGGATTTAAAAAATACTATATGCTGTCTACAAGAAGT
ACAAGGAGCCCACTCCAAATTTGTAGACACACATAGGATAAAATTAAAAGGATGGAAGAA
AGTATTCCATGTGAATGGTAACCAGATGAGAGCAGGGCTCATTATACTTATATCGGACAA
ATAAATTGTAAGTCAATAATTGTCACAAGGAACAAAGAAGGACAATATGTAATATTAAAA
GAGTCAATTCACCAGAAAGATATAACAATTTTAAACATATATGTATTCAATCTTAGGGCT
TTAAAATATATAAACAAATATTAATGGAACTGAAGGGAGAAAGACAGCAATACAACAATA
GTAGGAGATTTTAATTCTCAGCTTTCTTTTTCTAGAGACAGAGTCTCACTCTGTCACTCA
GGCTGGAGGGCAATGGTACAATCTCAGCTCACTGCAATCTCCACTTCCCAGACTCAAGTG
ATTCTCCCACTTCAGCCTGCTGAGTAGCTGGGACTGCAGACATGCAACACCATACCCAGC
TAATTTTTTAACTTTTTGTACAGATGAAGTCTCGTATATTGCCCAGCTGGTCTTAAACTC
TTGGGCTCAAGTGATCCTTCACCTGGGCCTCCCAAAGTGCTGGGATTATAGGCATGAGCC
ACCGTGCTCAGGACCCAACTTTCAAAAATTGATAGAACATCCAGACAGAAGATCAATGAG
AAGCGGATTGAACAACGTAGACCAAATAAGCCTAACAAACATATGCAGAAAATTCCATCT
AACAGCACCAGAATATGCATTCTTCTAATGCACACACACATATTATCCAGAATAGATCAT
ATGCTGTGTCACAAAACATGTTTTAACAAATTTAAAAATACAGAAATCATATCAAATATC
TTTTCTGAACACAGTGGAATGAAACTATAAATCAATTATAAAAGGAAACTGGCAATTTCA
CCAATATGTGTACATTAAACAATAAATTCTTGAACAGTCCATGAGTCAAAGAAGAAATTA
TAAGGGATATTTGAAATGTTTCAAGATAAATGAAAATGTCTCAAGATGAAATAAAAAGAC
AACATATCCAAATTTATGGAATGCAACAAAAGTGGCAAGAGTTAAGTTTATAGTGGTAAG
TGACTACATTATAAAAGAAAAAAGATTTTAAGTAAACAACCTAACTTTACACCTCAGAAG
TGGAAGAAGGAGAAAATACTAAGCCTAATGTTAGCAAAGAAAGGAAATAATAAAAATTAG
AAAAAATAAATTAAATAGAAAGTAGAAAATTACTATAATAATTAATGAAACTAACAGCTG
CTTTTTAAAGATCAATAAAATTTACAAACCTTTGGCTAGAATAACTAAGAAAAAAGAGAG
AAGACTCATAAATAATATTGTAAATAAAAAAGGAGCTATTGCAATCAAAGAGGCAGGAAC
AATAAAGATTTTCAGGCTATTCTGTATAATTATACACTAACAAATTGGATAACCTAGAAG
AAATGTATAAATTCTCAGAAATACACAACCTACCAAGACTGAATCAAGAAGAAATACAGA
ATCTGAACAGATCTGTAACTAGTAAGGAGATTAAATCAATGATCAGAAACTTCCCAAAAA
AGAAAATCCCAGGATCAGAAAACTTCACTGGAGAATTCTGCCAACATTTAATAGAAAAAA
AAATGCCAATTCTTCTCAAACTTTTGCAAAAAATTGAAGAGGACGAAGCATTTCAAACTC
ATTTTATGAGTCCAGCATTTTCCTGATACCAAAATGAGATAAAGATATTACAACGAACAC
ACACACTTTCAAACAAGCTACAGGCCACTATCTCTGATGAATGTAAATGCAAAAGTTGTC
AATAAAAAATAGCAAACTGAATTCAACAGTGCATTAAAAGGATCACACACTGTGACCAAG
TTGAATTTATCTCTGGAATGATGAATGGTTTAACATATGAATATCAATCAATGTGATACA
CTATATTAACAGAACAAGGGATAAGATCACATGATAATCTCTATAAATGCTGAACAATCA
TTTGACAAAGTTTAATACCCTTTCGTAATAAAAATACTCAACAAACTATGAATAGAAGGC
ATGTACCTCAACACAATAATAAAGGTCACATATCAAAAGCTAACAGATAACATCATACTC
AATGGTAAAAACTGAAAGCTTTTCCTCCAAGATCAGGAACTAGGTAAGAATGTCCATTCT
TGCCATTTCTCATCAACGTATTACTAGAAGTCTTTGCTAGAACAATTATGCAAGAATAAG
AAATAAAAAGCACTGAAATCAGCAAGGAAGAGGGAAAATTATCTCTATTCCCAGATATAA
TAATCTTATATGTAGAAAATTCTAAAAATCACACAAGGAAACTGTTGCAACTAGTAAGTT

FIG. 1U

CATCAAAATTGCAGAACATAAAATCGAAATGCAAAAATCAGTTATGTTTCTATACAATAG
CAGCAAACTCTCTGAAAAAGACATTACAATCCCACTTACAATATTATCAAAAATGACTAA
AATGTTTAGTAATAAGCTTAACCAAGGAGGCTAACGACTTATACACTGAAAACCATAAAA
GCATTACCAAAAAATAATTTTAAAAGACACAAATAAATAGAAAGATAATTCTGTTTTCAT
GGGTTAGAAAACTCGATATTGTTAAAATGTGCACACTGCTGAAAGCAATTTATAGATCCT
ATACAATCTTACCAAAATTATGATGTCATTTTTTTCAGAAATAGAAAAAAAATCTGAGAA
CCATGGATACTTAGAAAATCTGGAGAAAGAAGAGCAAAGTAGAGGGTCTCATGCTTCCTG
ACTTCAAAACATATTCCAAAGCCATTGTAATAGAAACAGTTTAGCACTGGCATAAAGACA
GATATATGAACTTACAAACCAGCATAGCGAGCCCAGAAATAAGCCCACACATACATTGTA
AAATAATATACAAAGCACAAAGACTATGGACAGGATAGTCTCTTCAACAATTGTGTTGGG
AAAACTAGATAGCCATATTCAAAGGACTGAAATTAGACCCTACTCAAAAAATCAAGTCAA
AATGAATTAAAAATTAAAGATCTGGGCCGGGCGTGGTGGCTCACGCCTGTAATCCCAGCA
CTTTGGGAGGCCAAGGGGGTCAGATCACGAGGTCAGGAGATCGAGACCATCCTGGCTAAC
ACAGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGCGTGGTGGTGGGCGC
CTGTAGTCCCAACTACTCAGGAGGCTGAGGCAGGAGAATGGCGTGAACCTCAGAGGCAGA
GCTTGCAGTGAGGTGAGATCACGCCACTGCACTCCAGCCTGGGGGACAGAGCAAGACTCC
ATCTCAAAAAAAAAAAAAAAATACAAGATCTGAAACTATGAAACTCATAGAGAAAAAACAG
GAGAAAAGTTTTATACCATTGGTTTTGGCAATAATTTCTTGTATACGACACCAAAGAACA
GGCAGTAAAAGCAACAAAAAATAGATAAGTGGAACTACATAAAATTAAAAACTGATGCAC
AGAAAATAAATAAAAAGAAAAAACAGAGTGTAAAAGCAAACCATGAAATGGGAGAGAATA
TTTGCAAACCATATATCTGATAATGGGTTAGTATTCAAAATATATAAGGAACACCTACAA
CTCAATAGCAAAAAACTAACCCAATTAAAAATGGACAATGGACCTGATGGATATCTCTCC
AAAGAAGATGTAAAAACAGCCAACAGATACATGAAGAGTGCTTAACATCATTAGTAATTA
GGGAAATGCAAACCAAACCACATGAGCTATCATCTTACACCTGGTAGGATGACCATTATG
AAACAAAAGAAAGAGAATTAAAAAAAAAAAAGTGTTGAAAGGGATGTGGAGAAACTAGAA
CCTTTGTACAGCCACTGTGAAAAAATGTTTGGAAGTTCCTCAAAAAAATTAAAAATAAAA
CTATACGATCCAGTAATCCCACTTTTAGATACTTTTCCAAAATATTTGAAAACAGGAACT
CAAAGAGATATTTGCACTCTCATGTTTATTGTAGCCTTATTTACAATAGTCAAGAGGTGG
AAACAAATGAAATATATAATGACAGATGAGTCAATAAAATGTGGCATGTACATATCATGG
AATATTATTCAGCATTACAAAAGAAGAAAATCTTATAATATGCTGCAACATAGACAAACC
TTGAGGACCTTATACTAAATAAAATAAACCAGTCACAGAATGACAAATACTGCATGAATA
TACTTCTATGAAGTATCTAAAGTAGTCAGTCATAGAAGCAGGAAGCAGAACGGCAGCTGC
CAGGTCCTGGGAGTAAGAGTAAGAGGAAAGTTGCATTTCAGTGGGTATAGAGTTTAAAGC
ATGCAAGATGAAAAAGCTCTAAAGATCTGATGTACAATAATATGCATATAATGAACAATA
TTGTACTGTTCACTTAAATATGTGTTAGGTCCATGTTATGTGATTTTTACCACATTTTTT
TGAAAGCAAGTTGCTAAAGAATTTGCCAAATGGAATTATAGTGACACGAGTTCAAATAAA
ATTAAAAAACGAGAAACAGTAGAGTTTACTTAATTTGTTAATATATCCATATTATCATTT
TAGGGAATTTTTACTAAAGCAGAGTATATAAACTATCTTTTTTTGTTCTAATGATCCATT
TGTTTTAGTTTGTTTCCCATTTTTATGTAGCTAGACTGCCAGTTAATCTCCTAAAATTAT
TGGCACCATATTTCCCATTTTTTCTGGCTTTTTTATTAGTAACTGGGATCCTTGCAGCTG
TATCTATGTGATGCCAAACAATTAGGTTGATCAATTCTGTGACAACAAGCCATCTGGTTA
CTTTAGTGAATAGGCCCTTACTTACCTTTCATAAGTTGATTCTATTCTCCTTTGTGCCTT
CTCTTTAAATTACCATTATCCTGTAACCATAAATTAAAAATACAGCATCGCTTTTAAAAC
ATCCTGAAGTAATTTTTAACACTACAAAAGAGAAGAAATTTCCTTTGTTTGGTGTTCTTT
GACCCTAATTAGCATTTAGGAACAAACTACACTTGCAAAATTATTTTCGATTGGTAGAGG
GAAGAAAAGGGTCTTTTTATTACTATGTATTTGTAATTACTTTTGTCACTTATGTTATTC
TTGTGTCTAAATTCAACTCTAGATTTATTCTCTGTTGATATTTTTTATCACTTGAGAATA
TTTTAGTTTTTCAACCTCTATATGGCGGGCTATCACTCCAAATTTAGGTTAAACTGTAGG
TTGATTTAAAAATCTGGCTATGATGCAGAAAAAATTCGGGCAACTTACCTAGAAAAAAAA
AGTAGTTATATTTCAGTACTTCTTTTACCTAATCAGCCATTTTAAAATAATTTTGTTCAT
TATCAATATGGAGGAAATTATTTATATGCAGGGAAGTTATTTATATGCAGAGCTGTTAAT
GGCAGCAATCTGCATGACAAATTTCTACTTAATAAGCAATGAAATAGTTGGATAAATGTG
TATTTCTACATGGGTGAATTTCCCAAAATTCACACTTCAAAGACAGTTGCTGACATTTTT
TCAATGAGAGATTTTATTAGATAATGAGTCATCTTAGAGTTATCTTGTAAGTATTCTTTA
GTCTTAATTTAAATTTAAATGAAAGTCAATTCAAAGTGTTGTATTTTCTTAAATAAATTT
TGTTTTATAAACATTAGAAATTAAATAGGACTACCATATGGTCTAGCAATCACACTTCTG
GGTATATATCCAAAGAAAATCAGTTCAGTATGTCAAAGAGATGTTTCGTATTCATTGCAG

FIG. 1V

CTTTATTCACAATAGCCAAGATATAGAATCAATCTAAGTGCCCATCAATGGATAAACGTA
GAAAACATGGGCTGGGTGCGGTGGCTCACGCCTGTAATCGCAGCACTTTGGGAGGCCGAG
GCGGGCAGATCACGAGATCAGGAGATCCAGACCATCCTGGCTAACACGGTGAAACCCCAT
CTCCACTAAAAAAAAATACAAAAAAAATTAGCCGGGCATGGTGGTGGGCGCCTGTAGTCC
CAGCTACCCGGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAGGCGGAGCTTGCAGT
GAGCCGAGGTTGTGCCACTGAACTCCAGCCTGGGCTACAGAACGAGACTCCGTCTCAGTT
AAAAAAAAAAAAAGGAAAGAAAACGTGGTATATATACACAATGGAATACTATTTAGCCTT
TTAAAAGAAGGAAACCCTGTCATTTGCAACAACATGGATGAACCTGAAAAACATGTTAAG
AGGAACAAGTCAGGCACAAATACTTAATGATCTCGCTTATATGTGAAATCTAAAAAAGTT
GACTTCATGGAAATATAGAGTAGAATGGTGATTATCGGGTGCTGGGAGTTGGGGTAAGAT
GTGGTTGGGGAAACGGTCAAAGAATAAAAAATTTCAGTTAAAGAGGAAGAATACATTCAA
GAGATCTATTGTACATGTTGAATATAGTTAGTAACAATATTTTGTATCCTCAAATTGCTA
AGAGAGTAGATTTTAAGTGTTTTTGACACAAAAACTGATAATTATGTGAGGTAATACATT
TTTTAATTAGCTCCCTTTAGCCATTCCACAATGTATACATCTTTTAAAACATCATGTTGT
ACATGACAAATATATACAATTTTTATTTGTCAACTTAAAAAATATTAAAGATTTAATGTA
GATAAATGAAAGAAAATTAGGAATTAAGGTACAAAAATTATTTATAGTGTTTATTATTGG
TCTATGTTTACATAGTATTTCTTTGTCTCCATTAGTGTGTTATACAAATACCCAACTAGA
AACATGACTTTACAAATGGTGTATCTGATCTTTTATGTCCCTAGTTATTATTTTAGCCCT
GTCTTTTTTTTTAATAAAACATATTCTGCTTTTTCTTGTCCTCATCCTTCTATGAGTTGA
ATTAGTGACTCTACTCCAAAGTAATGGTGTTGCTTTCTCAGACCATATGGTGATACAAAG
GCATATGAGTTATCATAAGCATGGTCTGTGTAGGCAAAGCATGTAACTCCACAAATGCTT
CTTGAGAGATTCTAATATAATCTGTGCCAGACCTGCACAAGGCATAGAGAATAAAAATTT
GCACCCCACACAGTCACTCCTCATTCATTCATTCAACAATAATCAAGTACCTGGTAATGC
TAATGCAGTGTACTATAATTCCATATACATAAACTAATATTTTTAAGATACATGAAGGTT
ATGTTATAACTAATAGTCAATGTATTTTTAAAATTACTGTAATCAAATTGTAATTGTAAT
TAAGTATTTTCTTAATCAACAGAAACTAAAAGTATAATTTCCATCAACTCCTTTTAAGTA
TAAATGTAATTAAATGCCTGGCACATTCTTCACATTATATAAGGATCTTTATACTTAAGA
CATTTGGGAAACCCTACTTAGGCTTATCATTGACAAAACATTTTCAAAATCTTTTCATTT
GGTCCTCACCACAATACTGTTAAAAAGACAGCCTAAGCTGTTTTGTGCTTCCTCCCTAGT
TGGGCATCCCTGTGCAATGAGAGGGACAAACAAGGTGGTTTTAAGGTCAGAAACATCCAA
TTGCAGCATCATTGGGAAATTTGTAAGAGCAGCTTTTATAAAATGTCACCAACTCATGTA
TCTTTAAAAGATGTGCTGAATCTTATGCCTTGAGATTTTTCTTAGTTTCCTTATTTTCTA
TTCCCCTCCCACTTTCTCTTTGTCCCTTGGTGGCTTCATTAATCCCATATTACAATACAA
AGTAAATAATAGTGCTCTGAAGTGCTTCCTATTTGTTCAGGATGAAGTCTGAAAAATGAA
ACTGCAATTTTTTTTCTTTTGAGACAAAGTCTCACTCTGTTGCCCAGGCTGGAGTGCAAT
GGTACCATTTCAGCTCACTGCAACCTCCGACTCCCAAGTTCAAGTGATTCTCCTGCCTCA
TCCTCCCCAGTACCTGGGATTACAGGCATGCACCACCACGCCTGGCTAATTTTTGTATTT
TTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGGTGGTCTCGAGCTCCTAACCTCAGAT
GATCTGCACACCTTGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACTGAGCCCTG
CCAAAAACTGCAATTTTATCTTAGGGGACAGGTAAGCATAAAAACATCCAAAATCATGTA
TTTATGTTTAGGCTCTGCTTGTAGAGTGATACCAAATTCCAGGTGTTTTTTTTTTTTTTT
TTTTTTTTTGAGACAGAGTCTGGCTCTGTCGCCCAGGCCTGGAGTGCAGTGGTGAGATCT
CGGCTCACTGAAAGCTCCGCCTCCCGGGTTCACACCATTCTCCTGCCTCAGCCTCCCGAG
TAGCTGGGACTACAGGTGCCCGCCACCACGCCCGGCTAATTGTGATTCTTTACATTATCA
AAGAATTCATGAAAACAGGATATGAAGATTAGTGAAGGATTCTTTTCATTAGCAAAGTAA
CTTTTCTTATTTCAAATTTAACACATCTATTTATAAAAGTTATAGAATTTAAATTTTAAA
ATATGAATGAAGAAAAACAAAATCAGCATAACATAGTAATACATATAATTGATATGTACT
ATTCTGTTACTTGGATTCATTACTTAACCCTTGCAGTATTCTATGATTTTTTTTAATCC
ATGTGTTACAGTTAGGGCTTAGAAAGATTTAAGCACCTAGCCAAAATTATGCATTATGTT
AAGTGGTTGATATCCACTTATTGACAAATATGTATTGATTGAGAATTAGTCATGGAGATA
TCAATGGGTTATTTGATTACTTTTTCCATTACTCCCAAGTGGTCAGGATTAGTTTTAGA
TTATTTAAGTAGGTTGGCTGAGTTCACAAAAGCTATTACTATGGGGACCTTAATTGAAAT
CTAACTCTATCCAATTCTATTTCTTTTCCCTATCCCTCGAATGGGTGTATGTGTGTGTGT
GTGTGTTTGCACACATAAAAACCTGTTCTAATTTTATGCAACATGGAAAGCATTAATGTT
TAACATGTATGTTTGAACAGGGAATTTTGTACTGCATTAAAGATTATTCCTGTGTATTAC
ATACAATCAAATATTTGACTATTGACTGTCTTAGTATGTTCATCTAATTGTTTCCTATTC
CCATGAAAACTGTATCAGTCTGAGAACAGCTACTATATGATATGCATCACTAGTCTCCCC

FIG. 1W

ATGGTGCATAATACTTGATATAAATTAGATGCTGTTGGTTATACTTGGCGGGGGGAAAGG
GGACACTAAAAAGGAAGAGTCAATTTCTACTGTGAACAAAGCAAAAAGCAAAAGGAGAGA
TAAATGGAATTAAATTAAAAATGAAATTGAGAGTGTAGATAAATCTATGTAATGAAGATG
CTAGTAACATAGGAAGAGAAATAAGATAGGGTATAACAGTGATTATTTTTCCTAATAAGT
AGTGTCATGGCAGTTGGAAGACAAGAGATTATCCAAGCACTGGTTATAGTCTGAAAGATG
AGGTGGTAGCTTACTTGTTTGGGCCTCAGGCATTGCAGTACAAACAGACAGTGAGGGAGG
AGTCAATTAAGACTTATACAAATGCAGAAGTCATGGTTGAGGTAGTGAGAGGATTTCCAG
GACAGTGATGAATAACAGAACCTCAGCAGAAGGAGCATGTGGACCCAAAGCATCATACGA
ATAATGATAGGACCAAGGGAAAAGAAGTCAAGCGGAATGGGGATAGACAAAAGTTTTGAA
ATTTATGTGTAAGAGTTGAATGAAGAAAGTTATTAATAAGACTTACACAACAAAGAATTT
CTACATAGAAGTTGAAAAGACAGCAACAGAGTTTAGAGTTTAGGAAAAAAAAATTAAATAT
TAAATTTTAATATGTAATATTGTAGGATTTGAATACCTTAAAGCTGAAATTCAGTTTTTG
ATGCTGCTTCTTAGCATCTTTGTCTTGACATGTATATCAAAATGTAAGAATGTCTGTATC
TTACAATCTGTGATTCTTGAGAAGTCAATGCCATATTATTCACTACATTCATTCTTTCTT
ATTGGAACCATAATACTTTCTTCAATAATAATGTCAGTAGACATTCTAAATAAATAAAAA
ATATCCAATAACATGCCCCAATGTTTCACAGGTATCACACCAATAGCCCCTGAGATATTG
TCACATTCCATTTATCTGCAGAAGTCTTATTCAACTTTCTGTATTAAGTACCAAGAAATT
TCTTAGGCAATTAGTAAGTTCACTTGTATTCTTAAAACTTCACAGAATGAAAAATTAAAA
ATTTTAATCTCTTTTTCTAGAACAATTGTTTTACAAAGACTTTTCAAGGTTTTTTAATCC
TATTTTTTGACAAAATAACATATTTTAATGAAAGTAAACATGTAGAAATGACTTAACCAA
AACTAGCTATTGACAACTTTTCAGCACTTTTTTTGGGTGAATTCAGGAACAAACTTTGT
ATTCATTTTATTAATCCACTAAGTAGGGTTGCTTCACTTCCTTGGTTACTGTGCATGTGG
ACGAGGCTGATTTTCATGGTGGGATGTTAAAAGGAGGGATTTTTGCAAATCAAACCACAG
AACCATCACCTCACACTTGTTAGGATAACAAACATTAGCAAAACCAAAGATGACAAATGC
TAGCAAGGATGTGGAGAAATTGGAACTCCTGTATATGCTGACAGAAATATAAAATGATGC
AGCCACTATAAAAATTTTTTGTTTTTGAGAATGTGTCTTGCTATGTTGTCCAAGCTGGCA
TCAAACTCCAAGACTCAAGTGATCCTTTCACCTCAGCCTCCTGAAGAGCTGGAACTATAG
GCATGAACCACTGTGCTGGCTTGGAATATTTTTATTTTCCTCAAAAAATCAAAAATAGAA
TCACCATATGAGCCAGCAATTCCATTTTTGGGTATATATCCAAAATAATTTAAATCAAAA
TGTTGAAGAGATATCTGCACTCTCACATTCATTGCAGTAGTCTTCACAAAACAACCTAAA
TGTCCATCCATGGATTAATGGGTAAAGAAAATATGGTCTACACATACAATGGAATATTAT
TCAGCCTTAAAAAAGAAGGGTATCTTTCTGAATGCAACATCATAGATGAACCTGCAGGAC
GTTATGCTAGGTGGAATAAGCCAGGTATAGAAGGACAATTATTGCATGATTCTACTTACA
TTAGGTATTTGAAATAGTCAAACTCATGGAAACAGAGACTAGAATGGTAGTTGCCAGGGG
CTGGGAGGAGGCAGAAATGAGGAACTGCTGTCCAATGAGTATGTAGTTTGAATTATGAAA
AAATGAATAGGTTCTAGAGATCTGCTGTACAACATTGTGCCTACAGTTAATGATGCAGTA
TTATGCACTTAAACATTTATCAAGAGAGGAGATGCCATGTTGAGTGCTCTTTTCACAATG
AAAGTACAGTAAAATGAAATGAAATATACAGCAGGCTTTACACACACCGCTTCACAGGCA
AAAACTACTTGGGAAACAAAATGGAAGGTCCCCAGAGTCGTGAGGGAAGTAAGGTATGGT
ACAGGGTCAAAATGGCTGTACCTGGAGCTCTCTGACTGGTCAGGCACCAACCAGCAATAC
TCTCATGCCTTAATTATAGTTTACTGCTGAGATAATTGAGAATGAGAGCTCATATTTACT
AACCAGGATATGAATAGACTGAGAACTTTAAATAACTTTCCTTTAATTCCATAAAAATCT
CCATTCTGTTTTAAAGTCTTTAGTACAGATTTTAGATGTAATAAACTGCTAAGATTTGAG
CAACAACTATAAGCATAATAAATGGTTTGCTTTATGGGCAGTTTTACACTAATGCCTCTA
ATAATAATAACAGTAGCAATAACAAAAATGACAGGATTCTTAGGACTTCATTACTCAGAG
CATAATCCCTAGAAAGCAGCAGTCATTATCTAACCCAGAAACTCCCAAGAGTTTGCTTAA
CACTTTAAAATGTATAATCTAAATTAAAGAAAATATGAGTAAATGGTATTGTTTCCCCTG
AATTGAAGTAATATGGGATGTGTTGAAAGAATACATCAAGACATTTTTCACTGTCACCTA
GCCTGATGACTGACATAGATTAATTACTACATAAATTTCCTCTTCCATTTAATACTGATA
AACAGATTTATGGGACTTAAACCACAGTACACAGTTTTGTATTTTGTACGAAATGGATAA
TCACATTTTAAAACATGTGTAAGGCATATTTGCAAACTTGAAACGTCGTCTTCCATAAAT
ATATGCTGAATGAATGAATTAATGAATAAAAATTGAGGCAAAAACTCAGGTGTGGCTCAG
TCATCTGAATGTTATTATCCAATGAAACAGGTCAAAGATTTTTTTTTTTTTTTACGGTTC
ATTTCTAGCCAATAAGACCAAGGTTCATTCACTTCACCTCTGTATAGAATCCTTTGTTGG
GGGCTGCGAGGAGGCAGTAAGAAGTATCACATCTAATCTTTTCCATAATTAGCCAAGTTA
GTTGGTACTTCCCATAACTCTGATACCCATAGGCCCTTGCTATTTCTAGACTTGAGTGTC
ATTCAGAAATATGGTTTAGGCGAGCACTAGGAAAGATACACAGTTTTTCTAAAACACATT

FIG. 1X

ATCCAATCAATATTCTACTTATAAAAGTCAACTACACACACTTCAGTCATGAGGTAAAAA
AATGAAATTTATACATAACACTCACTTATGTTTATCACTCACTTATATTTATAATAATAG
ACATACAGGTATTCTATTAAAGGAACTTTTTAATGTTTGACCAGAAAAAAATTTCAATATC
CCTTTTTATTAAGTTTAAGTTACTGTAATGAAATTAAACATGTGAAGGGAGACTAATACT
CTCTTTTAAGAGAAGTAAGAATGAAATATCCATATAAAATACACTGCATTATTCTCTTTG
TTTCAATGGCAAATAGAATCAAAAGGAATAACCCACTTTATTTAACGGAATATCTGAAAG
TGTTCCACTTATTTATTTCTAATTTTAACTATGGAAAGTACTTGCATTTTTTTTTAGGAA
AGAAAGCCAAGATTTTATAAAGTAAAAATCTGCTTTGTGTGCCTTTCCAAATTAGAAGAG
AAATGTATCATCTTAATACAGCAGATTCAGTTATTATAAAGACCTACTCCATCCAAAAAA
TTGAGTGAAATAAAAAGAAATTGACTTACTTGTTAAAGAGAAAAGATTGCCAAGGCTTGC
AGACTTGTGAGGTGGTTAAATAACAAACTAAAGACTAGCGAATATGAGCTATTTTGTTTG
ACGTGCCTTCCATTTAATAAATGCTGTATCAATCTAGCTGTTTCTCTATTTTTAATCATA
CATTTTGTTGTTGCTCTAAATTTAATCTTACCTTATACATTGTATAATAGATGTCCCTTA
AATACATCAAATTTAACGTGTTCCAAAGAAAACTCATAATCTCCTCATCTCCATCCACCT
CACTCCTCCTGCTGTGATCAGTCTCTCCGTTTTTGTTCATTGTCCATCATCTTCTACAGA
ACAGATGTGTCCTAACCCACTTTCCTAAACACATTTTTGTATACAAAATAATTTCCTTTT
TTTAATTTCAGAACTCTATTCTGACAAACATTTGGCTTCAACCTGTAATTAAAAACTTAA
CAATACTTAATAGTTGCCTCAAAGAGCATCCCCTCTTTGTCAATGTGAGACTATTTACAT
TAATTTACATGTAATTCAGTTTCATACTCATTCACTGGGGTGTGAATATTAGTCAAACGG
GCAATTAATTAATACAATCTTTATATATTCACTTATTAAAATGCACCACACAATTCCTAA
TTTATTGAGAGTTCTCACTAAATCTATGGGATGTAAATTTTGAAACAGCTGCAGCTGTTT
ATGCCATTGCTCTTGTTGTCCAATAGAGCCAAGTGGACATTCTTTTTTGTTGTTGTTCTT
TCCTTGAATAGAGTCGAAATTATGAATCTAACTTTCTCCGACATGTTGTCTAAAAGGATA
TCATCTTACCTTACTCAGTGTGAGCCCTAAAACTAGGAAATGTTTATCAATCTCTGATTG
CAGATCAAGTTTAACTATCAAATACAGATTAACTTTTCAGCAAAAATTTGTTAAATATTC
AGAGATAGAAATCTTGATGTTGGATGACAAAGATCACTTGTGAAGAACTTTATTAAGTTT
TATTTGGTTGAAAAATCTATAATTTTTAGTGAACAACTATCATCCATTATGTTCCAAGCT
TTGTGACAACTGTTTTTATGTCCATTAAAACAGTCCTATAAAATAGGTACAAGTATCTCA
ATCTTATACATGTCAAAACTAAAGCACAGAGATGCTAAATAACTTGACTAAACAAGATAT
TGAAGGTGAAGTCTGAGATAGATTTTTAACTCCGAAGTGCATAAACTTTACCTCTATATT
ATCTGTCTTCAAAAAGAATGATTTTAAAGATTAGGCTTTTTTATTTCAGAAGAAAATATT
TTTACACAATTCTAGATTCTTAACAGTAATTTGAAGGAATGAATGTCTGATGATTCAAGA
AAAGTGAGGTACATTTTAAAGGAAAAGTGACAGACAAAAAATGGATTTTTGAAAAATGAA
TAAAGCTGCTTTTTTTTTTTTTGATGGTGTCTTGCTCTGTTGCTCACGCTGGAGTGCAATG
GTGCAATCTCAGCTCACTGCAATCTCCGCCTCTCGGATTCTAGTGATTCTCCTGCCTCGG
CATCCCGAGTAGCTGGGATTACAGGCGCCCACCACCAGACTCAGCTAATTTTCTGTATTT
TTTAGTAAACATGGGGTTTTACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGG
TGATCCACCCACCTCGGCTTCCCAAAGTGCTGGGATTACCGGCATGAGCCACCACGCATG
GCCAAAGCTGGTTTTTAAAAGGGATCATTGTACATTATTATCAAATTTCATTTGAACGTC
AAAAATTCTGAGGCAAGAAGGAAATTGAGCCCAGGAGTTTGAGACCAGCCTGGACAAAAT
GGCAAGACCCCATCTTTACAAAAACAAAAATAAAATAACACTAGCCAGGCATGGTGGTGC
ACACCTATAGTTGTAGCTACTTGGGAAGCTGAGGTGGAAGGATTACTTGAGTACAGAGAA
GAGGTTACAATGAGGGAGGATCGTGCCACTGCACTCTAGCCTGGGCAAAAGAGCAAGACC
CTGTCTCTAAAGAATAACAAATAAATAAATAAAGTCTGGACAAGCCTAAAATCAGTAATA
TTTGGGGAATATGCAAATAGTCTTTGCTTTATTTACTCAATTATTGAAACTATATTCAAA
AATAGGAAGTAAAACATGATTTAATATTATTTAGTAAGTTAAACATGTTATAATAATTTG
GAAATCCATGTATGTTAGTTAAATATACATTACTATAAAATGTAAATCAGTGTGGTTTGT
AGCAGAGACCTGGATTTTTTATCTTTGTAGTGTACCTACACCATCACAGAAAGGTTTGCC
ATCAGTCTCTAGATTAGGTGCAAATTCATTTAATGTGATCCATCCTATTATCTAAAAGGT
CATTCTGTTGTTTTCAGCCTTCATCTAAGACACTCTCAGATACTATTTCAGGAATTTATG
ACAGCAAAATGATATAAGGTGACAAAGTAGAAATAGGTGCTATGCTGCTTTACCTATATT
GAGTTATTTTCTTCTCTCCAGGATCAGATATTAATGATAAATTCTCTAACATCAAAAAAT
AAAACCTAGGTCATATAAATTTTACACAATCAATGTCAGTCACTCAGCAACCATTGAGAA
TCTACTATGTTTAGAATGGAACACCTGACTTATAGAAAAAAAGGTAAAAGATTGGTTTTG
TAAAATGACACATACAATTTAAGAAAAAATAGGCTATCTATATTAGATAGTTAAAAGAAG
ATTTTAAAATACGATAAGAAGAGAGGGGAGAAATGGCTAGATTAATTTGAGGATTACCTA
GTGTTAAAATAAGTCCAGATTTAAATCAAGTTTATTAATTCTGAAAAAGATCACATCCTA

FIG. 1Y

AAGAAGGCATCAAATTGACCCATAAATGTGGATAAAACTTCTGTAAGATAATGAAAGCCC
TAGAGAGTAATGTTCAACTCCATTTTCTAATTGGCAACAAATGTATAATATGGGTACACC
AGAATATCTAACTCAAAAAGTGGGGAAAAAAACTCAAAAAGTACGAAATGTTGGCAAAAA
TGCAGACAGCTAGGACACTCATACCAGCTGGTAAGTGTAAAAACTAGTACTGCACCAAGC
ACTTTAGAAAACTTAACGGCAGTTATGTAGTAATGGTGATCATATGCATACTCTATGATA
GCAATTTCACTGTTAGATATATAACTAACAGAAATTTGCACATATGTGTCAGAAGACGTA
CATAAGAATGTTAGTAACAGCCCTGTTTACAATAGCCCTGAATTAGAATGAACCAAAATT
TCCATCAATTGTAGAGTATTTCAATGATAATATAATCACACACTGGAATGAAAATGATGG
AACTACTACTAAACATACAACCTGGATCTTACAAACATAATCATAAGTGAAAGAAATTAG
ACACAAAATAACACATAAATGTTGATTCCACATAGATAAAGTTAAAAACAGATAACAATT
AATCTATGGTGTTACAAATCAGTATACGGATTTCCTTTTGTTGGCAGGGGGGATGTTGTT
GGAGAGGAAATAGGAAGAGAGCTTCTGGGGTGCCGGTCATATTGTACTTCTCAGTCTGAA
TAGTAGTTACAAGGGTATGTACACTCTGCTGTAATTTGTCCAGTGATACATGATGGTTTG
TACATTTTTATACATGTGTGATAATTCAATAAAAATATCTGAAAAGCTACAACAGCAGTG
GCAACAACAAAGCCCATTAACCACAAGAAATAATCATGTAAATTGTTTTCTTCAAATAAA
TGTGTTGTAAATAACTTCTCTCACTCTTTGGCATATATTTTTGTCCTCTTTTGATATACC
CTAATTTTAGGTTTGTTTAATTTTTCAAACATGTCCTTTATGTTTAATACATTTGAGGAA
ATCTGCTTAAGAAATGCTTATCTACTCCAACATCTTATCAATGGGAATTTTATTTTTTTA
ACTGTCAAATTTAGATCTATAAGTAACCTGGAATTTATGTTTGTATATGATGTGATGTAG
AAATCAAATTTTTATTTTTTCTATGTAGATATCAATTTATTCAGTATCATTTGTAGAAAA
GATACTTCTTTGATAATGCAGTACATGGCACTTTTGTCATATGTCAAGAGTCCTTATATA
CGTAGGTGTGGATCTCAACTATTTTTGTTTGTTTTGTTTTTGTTTTGGATCTCAATTTTT
ATTCTATTCCCTTGATCTACATTTATATCCTTGTACCAGTACTATACTGTTTTGTTTACT
GACACTTTGTATTAATATTTGATAGCTAATGTAAATCCTTCAAATTTGTTTTTCCATAAT
ATAATACTGACTAATTTTGGCCCATTATATTTTTATATAAATTTTGAAATCAGCTTGCCA
GTCTTTACCAAAGGAAAGCTAGCATTTTAATTTGGAATGCATTGAATCCATATATCAATT
TTAGAGAAAAACTCACAGCCTTACAATACTTATTCTTCGATTCCATGAGTAGGGTATATCC
CCCTATCCATTTAGGTTATTTTTCATATTCCTCATATTTTACAGTGCAGAAATCATGTGT
TTCTCATTATTTTTTTCCTAGATGTTGAACATTTATTATTCTATTGTCAATAGTATCATC
TATTTAAATTGCATTTTCTAGTTGTTTTTATTTAATAGAAACATAATTGATTTTGCATAT
ATACATTATATTTTATATATCAATTTCATGTGCTCTTATGTTACATATTGTTTTATATTC
AGCAAGTGTTACTAAGGTATTTATTAAGATTAGTAGTTTATCTGGAGATTCTTTCACACT
TAATAAGTATGCCCTCTGTGGATAATGATAGGTTTTATTTAATCCTTTCCAAACTTCATT
ATTTTATTTATTTTTATTGCTTTATTACCCTTGCTCCAGCACAATGCTAAATAGAAATTA
CCATAAAAGACTTTGTGCACTTACTCCTGATCACTGAGGGAAAGACTATTTATGTGAATT
AGTATTTGTAGATATTAACTTTTGAGAATTTAGCTGTCAATCCCAATATGACAACTTGGA
GGTGATGCATTTTTTTTCTTCCTGCTTTAAGATTTTTCCTTTCGTCACTGGTTTTTCAGCA
GTTTTATGATAATATAAGTGGGTGTGATTTTCTCTTATATTTATCCTGGTTGAAATTTAT
AGCACTTCTTATATCTACAAATATATACCTTTAATTCGTTTTGAAAAAATTCTTAGATAAT
GTATTTGCCTTGCCAATATCTTTTTAAAGATTGCTTTTGTCTCATGCTACTTCTATACAC
ACATATTGAGAATCCAATCACAGGTATAATAGAATTTTCACCATGTGTTATGCACACTCT
TCTGCATTTTCCTTTTTTCCTCTCTGTTCTTTAGCTTGGATATTTTCTATTAGTTTGTAT
AATCCTATTAGATGGTTTTATCTAATCTTTCTTTCTGTTAAATCTCTTTGTTGTGTTTCC
AGTTCACATATTTTTAAGTTCTATAATTTCCTTGGACTATTTTTCTATTTTTTATATTCT
TTATAATATATCTACTTTCTTGACATTATTAATTCAATCATTTTAAAATTTCTGAAATAT
TTTATGAAAAATTGTAGAAATTATTTTATGTTCTAGATAATATTATCTTCTTTCACAGAG
AATTTGCTTTTGCTTTGGCCAGCAGCTAGTGTTGGGACAGAAAACCACTATCCCGTCAGT
CACTGGAGGCTTTGGAAGCTGGGCTTCATTCTTTAGGAGAGCTTGTCTACTTCAGATTTA
TCCCTATCAGAGTTCATAACTTGGAGTTACAGCTGAAAGCCAGGGTTGTTTACCTACTTG
ATAGGCCTTGAACTCCAATTATCATCTTATTTTTGGTTAGGTACTAAATTTCCGGCTCAG
CATCTCATATTATCAGCTTTGTTCTCTGTTTCTCTTCTCCTGTTCTTAGCTAGAGTTTGC
AAATTGCCAAAAACTTTGAGAAGAAAAGAGGCTAAATGCCAGAGCATCTCCCTCTTGCAT
TTTCTCCAGGATATTGGCCTTTGATGTCCCTTCTGCCTTAGTAGCTTTCCAATGTCTTAA
AGAAATGTGTAACACTTCTGGTTGTTTTAGGTGGGAAGTTTGTTCTGCAGTAAGCTTATC
TGCCGTTACCAGAAATAGAAACTATTTTGTAATAGTAAAACAAATGTATACTTTCGTACT
ACAATATTTAGTACTTCAGAGAACAATTGGCACTTTCTGGATATTCTCAACCAGGAGTAT
GTGGTTGAAACTGCACAGTTTTCTGGAGATGATTTAGGTTCTTCCCTTCTTACTCTAATT

FIG. 1Z

CTGTCACTGGTTGATCTTATCCACTCCACAAGCTTTAATCACAATTTCTATTCTGATGAA
TCCCAAATATTTACATGTAAAGAAATTATATCCCCTGGAGTATAGAACCATAAATCTAAA
TGCCAACTGGGTATTGACACTAGGATAACTCACAGGTGCTTCAAAATTACATATACAAAG
TTGAATTTCTCATCTTCTATCTACTCTTACAAAGCTACCTCATTATCCTTTATCCCCTAG
CTCAGTGAGCATCCCCAGCTGTCAAGCAATATACCTGCTAATCATCCTCAGTTCTTCTTA
CTCTCTCATCCTCATATCTAATCCCTCACTAAGGCCTGATATTTCAACCTCGTTATTATT
TTTGGCATTCACCTTTTTCCATTTTTTGGTTACCAACTTGCTTTCTTGGAATTTTAAAAC
TGTCAGTATTAATCTCTCTGCTTGCAACATAAAGACATATATTTTCCACATATTCCGCCT
AAGTAATCTTTGAAAAATAGTAGTAAGATATTGCCATTCTGTTGCTTAAAATCTGTCAGT
AATTTTGTAATTTTCCAATTCTCCATAGTCTGTAGGACAATATCCAAATGTTTTAACTGA
ATACACACACACACAGAAACACACACACACGCTCACACACATTTTATGATTCATACTTTGAG
TTTAATTGAAAGATAGAACATCTATAAGATGAAAACAGTTGTAGTCAGAGATTCTGGTAT
GCAAAGTAGGAGAGAGAGCCAAGAACTAGAGGTATAACTTTGAATTATAATATTGGGTTG
GTCTTCTATAGATGAGACATAAAGTTGTGAGAGTCAATAAGAACAACTAAAGAAAGATAA
TGAAAGAACAAAAGACAAGTGGATTAAAGACAGATATGCGGTGAAAGAGAAAAGCATTTC
TACAGAAAAGACCCCCAAAATAAGTTCATTGCAGGTAGTAAGATGAACAGAAGTCAAATG
TCTTGGGGAGGATCGGATTGGTTGCTTGTGTATGTTAATTAATGCAAAAGGGTCAAAGAG
AAGGACTGACTTTATGGCCCTGTAGAACTCTGAGAACAGGGTCAAAATCCAGATGCATTT
CTAAGACATCACACTGGGAACGGGGACTTGTAATGAGTTATCTACAAAGTGTAAAAAGAT
GTGGGTAACCAAAAGGTTGTCATTTCCTCCAAAACAAATTTTCCTGGAGTGAAACTGTAA
CTACCAGGTATAGTCATTAATAGAACTGCAGACACTAAGACTATGGAACCTTCCGTCTTC
CTAACCTTCTCCTCAGGCCAGCCTTAAAGGCCTGTGAAGATCTATTAATAACACTGCTGT
TTTGTTCTCTGGCAGCTCTTGGTGCCAGAAGGCTTGGTGCCAATTTGTGGTTGAGCCCCT
CCTTGGGAGAAATCATGCCATTCAGAGACAGCTGATAAGTCAAGCCTATTTTCCCACTTT
CTTCACTGTATTTTTCCTGTCTGAAGAACTTGTTTATGGATTTGATTTCTGTAGAGATAA
TAATCACAGGATTCAGTGGTATAGCATTCCTCTATGCATTTTCTCCCTGCACATTTGTGT
GTGTGAAGATACTCTTTCTAAATCCCTTTCAAGACAAATTATTAATTGTGATATATTAAT
TATTCTCCACTGTACCTAACGGTATCAACACTACAGAGGCACCATTGGTTGACAAAAGT
GAGAGCTTTTCTCAACATTAACATAATGAGCAAGTGGCAATGAGAAAATATTTGTCCAAT
TAGAGACTTTTATATTTTCTTTTCTTGAGGAAATAAAACCCGAAACACATTTAAGATACA
TTGCTGTTTGTGCATAGGCGGTAAATTTTTTTTTTTTTTTTTTTTTTGAGACGGAGTCT
CACTCTGTCGCCCAGGCTGGAGCACAGTGGCACGATCTCGGCTCACTGCAACCCCCGCCT
CCCGGGTTCAAGCGATTCTCCCGCCTTAGCCTCCGGAGTAGCTGGGATTACAGGCGCATA
CCACCATGCCCAGCTAATTTTTGTATTTTTGTAGAGATGGGGTTTCGCCATGTTGGCCAG
GCCGGTCTTGAACACCTGACCGCGGGTGATCCCCCCGCCTCGTTCTCCCAAAGTGCCGGG
ATTACAGGTGTGAGCCACCGCGCCCGGCCAGTAAATAGTTTTGAAGTTTTATTTAATCCC
AGCACTTTGGGAGGCCGAGGCAGGGGGATCACGAGGTCAGAAGATCTAGACCATCCTGGC
TAACACCGTGAAACCCCGTCTCTACTAAAAATACAAAAAAATTAGCCAGGCGCGGTGGCG
GGCGCCTGTAGTTCCAGCTACTCAGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAG
GCGGAGCTTGCAGTGAGCCGAGATAGTGCCACTGCAGTTCGGCCTGGACGAAAGAGCGAG
ACTCCAGCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAGTTTTATTCATATTCATATT
AGATAACCATTTGGGTGGCACATTTCACAACACAGATGCACTTCTTAAGAGTCCTCCATC
CGTCAGCGTTGTAAAAAAGGAAGTGGCACGTTTGCATGTAGTTCTTCTGAGACGGAGATT
TAGGGACAACTTTGCCAAGGTGTGTAGGTGGAGAATGGGAGATTGAGACAGGCATATTGG
CTCAGGAAGACAAGGGAGTAAAACTAGCAATAGAAAGGAGGGCCAATGCCGTAACAGTGT
GATGGAGTGAAAACAAGAAAAAGGAAAATGCCTCAGGATTTGGTGGAGAGTTTGTTTTAC
CTTTTTAAGATAATACTCCTGGTCAGCTTCCCAGGTTCTTAAGTCTGGATACTGTAATGA
TTTTGGATGACTGCATTCCATGACCTGTTTCAAGGTAGGTTTTTTGAAAATAGGAGTTAA
ATATAGGCTTTCTTCCCTATGTATTCAGTTGCGTTTTTTTCTTTTTCATTTAGAAATGTT
GTTTTATTTCACGTTCTCTTATTTATATTTAATTGAGATGGTGTTGGCCATTTTATCCTT
CTTTTTTTTTGTTTTCTTTTCTTTTTTTATTTTATTATTATTATACTTTAAGTTTTATAG
TACATGTGCACAATGTGCAGGTTAGTTACATATGTATACATGTGCCATGCTGGTGTGCTG
CACCCATTAACTCGTCATTTAGCATTATGTATATCTCCCAATGCTATCCCTCCCCCTCCC
CCCACCCCACAACAGTCCCCAGAGTGTGATGTTCCCCTTCCTGTGTCCATGTGTTCTCAT
TGTTCAATTCCCACCTATGAGTGAGAATATGCGGTGTTTGGTTTTTTGTTCTTGCGATAG
TTTACTGACATTTTATCCTTCTTTAAACATTATTTTCTATCTAGAAAATCCAACTTCAAA
TAAATATACTCAGTTCTACATTATAAAAAGTATTACAATGAATTTAATGCTTAAAACTCA

FIG. 1AA

TTCCGGAAGTGACGATGGAAGCAGGTTCAAATGCTTTCACTGACACTTTGTGGCAAAGTG
TGGAACTACAGTATATTTTTCCAAGTTGTTTCCTGATATATTTTTTATGTACATAACAAT
CAATAAATTGTTATGCTATTTATTTATGTACTTATATGTAAATTAAACAACCAAGAAATC
GCAAAGTGTTTTATTAAGATGATATCTAAACTGAAATATCACAACTTACTACAAATAATA
CTTTGTTTCAAAAATAATTTGAATTGCATATAAAAATCACAGTTGCTGTGATTAACATTG
CATTGATATATTGGAACTAAGGTTTTTGGAAAAATTGTGTTTTCTTTCAATCTTTTAAAA
AATACCATATTTATAAAATGAGTCATTAAGATTATCCCTAGGCATTTTCATTCTGTATTG
AAGGTTTTTGAGGGACATCATTATTAGTTCAAAGTGTGTTTCACATTTTGTAGTCTGTCT
TACTATGGCAACTAATTTTTTTTTTTTTTTTTTTTTGTGAGAGGGAGCCTCACTCTG
TCGCCCAGGCTGGAGTGCAGTGGTGAAATCTCGGCTCACTGCAGCCTCCACCTCCCGGGT
TCAAGCGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGCTCCCACCACCAA
GCCCAGCTAATTTTTGTATTTTTTAGTAGAGACAGGATTTCACTATGTTGGCCAGTCTGG
TCTCGAACTCCTGATCTCAGGGGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTAC
AGGCATGAGCCACCACTCCCAGTCGGCAACTAATTTTTAAAATTGTGGTAAAATATACAT
AATATACAATTCAACAACTTAATCAGTTTTAAGTGTATAGTTCAATGACATTAAGTATAT
TCACCTTATAGTGCAACCATCGTCACTATCCACCTCCAGAACATTTAAAATTTTTTAAAA
CTGAAACTCTTCACTCATGGAACAATAATGCCTCCTTCCCCTCTTCTCCTAGCCCCTGGG
CAAAAAAAAAATCTACTTTCTATCTGTCTGATATGATTGCTCTGAGTACCTCATATAAGT
GGAATCATGTAATCATTGTCCCTCTCTGTTTTTACCTTATTTTAATATAATCAAAACTAA
ATAAATAAGCAAATTCTTAAAATAAAATTGATATATTTAGTACAGATCCTTTTGAGACAC
TCAGTGGTCCACTAATTATGTACCATATCCAATCACATCACAATATCATAAATTTTATAG
TCAATTATTAGTTGGCATTTCAAGGCCCAAGTATATGTTTAATAAGAGACACAATCTTAC
ATATGCAGTTTACATGTTTTTAATCTAGTCTTAGCACCAGCATATCACCTTAGTTTACAT
TTGTCTAAGTGCAAGTATTGGTTTTGGAATGTAATTTTGCTCATATACAATCTGTAAGAT
ACTAAAACAAAAGCTAGTTTATTATAAGTGAAATAATGGCAAAGGCCATTTTAAAAATAT
TGTATTATTTTCCCATTTGAAAATCAGTTTAGTCTTTAGCCCACAAAATAACAGGAAAAT
AACTTAAATCATAAAAACTATATCTGAATATTATTTAACATATTTTATAAAGATATCCTT
CTTTGGATCATGGCTGCAGATGTTTTCATGCAGCTTGAGCCACTTTCCATGTCTTACGGA
GAATGTGCAGGAGCTATATATCATCAGATTCTTTCAGAGAAAGAACCGGTAAGACAAATG
ACAGTCTGAAAGATAAAGGAAAAAAATAATTGATATCTTCTTGGCACCTCTGCATTTCAA
AAATACTATTTCAATAAAGTCCATGTTAGAGGTGGAATTCAAGAATTCACTGAATCTGCA
TTCTTGCCTTCTGCTATCCTCTTTTGCCCTCATTTGCTCAATTATTCCTCACTCCTGGTT
AATGAAGGCAGGCTTTTAAATACAGACTAACCATAAATTGACTTTAATATTGGTGTTTAA
TGGTTATTCACAGAACTGATTTAAAATGTGGTATCAAGTTCAGGTCCTGGGATTTACCAA
AGTTCATCAGAGGACACAGTACATGGCGAATTGAGAACCATAGCCTACTTTATGTCTAAG
AGAATATTGACAAACAGCTAAGTTCTCTGTGAGCTCTCAGATTTCACTCAAAAGAAATGA
AGAAAGTAAATTCTCTGTTTAGACTTTGTGCCTTTTTTCTCCTTTTAAAGAATTTGCTCA
TCGGAAAATATACCATACCAATGGCAGCAACATACTATAAGTTTATGAGCAAATCAATTC
CATCCATAGTTACTGCAGAATGTATTATAGGCAGTATTTTTGTTGGGAGAAAAGCAGCAG
AAACTTAGCAAAGTAAGGGAAAGAGAAAAAGCAGCTTATAATGATAAAGAGCCTTTGTGC
CCGTAGAGAGATAAGAAAAAATACAAAAGAAATCCATAATGATCCACAATAATTTTAGAA
TGCAATTTATGGCCATGAAGGGTACAACATGTGATTGGGTATCAAAGAAGAAAGAAGTCA
TGTTAATTTAGGCTAATTAAAAGATATTTTGTGAAGCAGAAAGTTTTTTATTTTGTTGGT
TGACCAGTTGATTTTGGACAGTTTTGGATACTATTTAATTGGTTAAAAAGCTATTGAAAT
GGAGTATCAACCATTTCCAGACAGAGGAATGGCATGAGTGATGGTCTGGGCACGGAATAT
GTTTGACACACAGTGAAATATCAGATTCACTCTGATGCTCTGTGTATTTTACGGGAAACA
TTATAAGGGATAAAGGGCAAAAATTCAACAGAAACCCAGTTACTATTGGCCATCTGAGAA
TTTTGTACTGTCCAGGAGAAAAGAGAGCTCTCATTGAAATGGAAGAGTTAATACAACAAG
ACATTGTGCTTGTCTGTACTCCTATATATTTTATCCATTAAAGGAATTAATGGATTTTAT
CCATTTTATGACATTTATTATTTTATGACACTTATCCATTAATGACATTAATGGATAAAA
CATATAGGAGTACAGACAGGCACAACGCATGGGGAAACTATTAGGAGGTCACTGCAATAC
TCTAGCTAATGGTTACAACAATCTGACATTGGGTATTTGCAATAGGAATAGAAAGAATAT
AATAGAGGAAAGAGATATTTTGGAGATTTCAAGCATAATTAATGGGAGAAAATGGAAGCT
TATACTTCAGAGAAGCACAAAGTCCAGTGATAAGTTTAAGTTGTATAAATTTAGTGTGCT
CTCAGGAGAAGGTGATGTTTACTTTGTACTTTTACAACCTTGCACGGGTGAGTGGGTTAC
TGAATAAACAAATAAATGTTTGTGTAACACAAATTTAGAGAATGTGCAGTTGTAGATATA
TATGTAGTTCTGAATAGTCCATTTAAAGACAGATACTAGGTTTTCTTCCAGGGTTTCTAG

FIG. 1AB

AGTTTCGGGTCTTACATTTAAGTCTTTAATCCATCTTCAGTTGCTATTTGTATATGGTGA
GAGATATGGGTTTAGTTTTGTTCTTCCGCATATGGCTAATCCAATTTTCCCAGCACCATT
TATTGAGTAAGGCGTCCTTCCCCAGTGTCTCTTTTTGTTGAGTTTGTTGAAGATAAATTG
CCTGTAGGTATGTGGTTTTATTTCTGGGTTTTCTATTATGTTCTATTGATCTATATGTCT
ATTTTTATACTATTAATAGTATCATGCTGTTTGGGTTACTATAGGCTTATAGCATAATTT
GAAGTCAGTTAATACGATACCCACAGCTTTGTTCATTTTGCTTAAGATTCATTTGACTAT
TTGGGCATAGCCACAGTCTTTAAATATTTGAATGGACATAATGTGAAAACCACACTTAAG
ATATGTTTAAACGGCACAGTAATATTATCTAACACAAACTCAAAATTCAAATGTATCCAG
TTGTCCCAATAGCTTTCTTTATAAATATCTTTTTTTCTTTTATTTCTTCTTAGGATTGA
AAGGTAAATATCCTAGCATCTACACAAGGGAACCGATGTGTGTGTGTATATATATATATA
TGTATATATATATACACACACACACACACATAGGAATACATACATGTATATATACCAGTATA
CACATAGAATACATAGGAAGATTTTTATATATATATATATATATATATATATATATATATA
TATATATATATATATATATCTTCCCCAAAAGTGTGCCTTGGCTTTTAAAAAAGCTTACAA
GATCTCAAACTGTCTTAATAGACTGACAGTAACCAAATCAATCATCCTTCTCATTGTTGC
TCTGAGTAGATTGCACCTGGAGAAATGATTGCAGGTATGGATAGCTCACTTAGAGCTATT
ACTGATAATCTGAAGTGTGTTCAGAATAAAATAACCAGGGTGATGGGGAATGAAAAGCCC
ATAAGTTTCACATGATGGATTCTGATTATCTTTAGGCTGGAGAAGCATAGGCTAGGGAAG
TGGGCATAGCTGTTGTTGTTAAATACTTGAATGAATGCCTTTTTGATTTGAATTGTGTTT
CTCCAAAAATATATGATTAAGTCCTAATGATCATTACTCAGAATGTGACCTTATTTGGAA
ATGGGGTCATTGCAGATGTAATTTGATATGGTAAAGTCATATTGCAGTAGGGTGGGCCTT
TAATCCAATATGACTGGGATCCTTATGAGATGATGGCCATGTGAAGATAGAAACACAGTA
GAATGTCATGCACTGACAAAGGCAGAAATTGGAGTTATACTGCACAAGCTAAAGAGCACC
AAAGATTGCCTGAAAACCACAAGAAAATAGGAAGAGACTAAGAAGAACTTTACTACAGCT
TTCAGAGACAGGACAGCCCTGCTGACACCTTGATTTGAGAGTTCTAGCTCCAGAACTGTG
AGACAATAAGTTTGTATTGTTTTAAGACACCAGGCTTATGGTACTTTTTTACAGCAGCCT
TAGAAAACAAATACAATGTACATATATAGGTAAAGCTTATTCATTCAGTTCCAAGAAATA
ATTAGGATCTTGTAAGCAGAAACGAAGGGAAAACAGAACATGAACAAGAACTTGCTAGTA
ATTAAAGCCACTGCAAAATGAACTCAAGGGCTCCAGCAGGTTTTAAATTACCTGGTATTA
TAAATGTTCAAGCAGGATGAATCAGAGATGGTGCAGAGGTGATTATTCATGCATCAGATG
GAAGGTTAGACTGAATAATCTCCAAGTGAAAAAATTATATGATCCTATCTTAAAGCCCTG
TCAAATAGAGGTTGGTAGCTTCCTTTTCATTTTCTGCTTCAATCAAGAGGATATATGGAT
GATATAGCTTGGTGGATAACACTTAAATTGAAGACCTAGTACTTAGTTTTACTTTTACTT
ACTCTAGTACTTAATTTTTCTTAGGTAGGCCCCTTAACTTCTCTCTCCTTATTTTCCCAC
CTGTTAAACAGAGATATTAATGCTATTCACTTCCTAGTGTTATTATGATGAAACTAGTTA
ATAATTTAAAAAATGCTTAGAACAAGGCACAGCACATAGTAATGACTAAAGAAAGAAGTG
CTTTTGAACATATATTTGCTCTACTATTGTCTAGATTGTCTAGATATAATGCATTAAGTC
TTCCCACCAGTGCCATTGCTCGTGTCCAAAATACAGAGTTAAAAGATTAGAAATAATTGC
ATGTTTTCTAAGAGTCCTGCGCATTTTCCTAGATCCAATATTGTACTATTTGGACAATTT
ATTGACCAAGTACCAGAAATATAATATTTTTGCCAATTTTCTCATAACAAACTGTGATAA
TGTGTATGTCAACTGCTAGGGTGGGTTTGTGTGTGTGTGAATATGTGTGTGTGTGTGTTT
CAAGTGTTTATAGAAAATAAATCACTCAATGGCATAATTTTCAAATAATAAAGACTACAG
TTACCCTGATTAAGGTTCACCTGAGTTTTGGATATTACCACGTGAGAGTTAGAGGACAAT
GTGAAGTTTTCAAAATTAAATCCTCTGAAATCCAGGTATCTTGTTAAATTGACATCTGTT
GGTAGCTGACAGCCAATTTCAGCTTCAGGAACTAGTAAGAACATTTTCCAGCTTATGAAA
CTATTAATAAATGTTACATAATTGTCCAAAGAAATCCTCATTCAGTGATTCAAATTTAAC
AAAATTAGGTTTTATTTATTCGCTATGTAAAGATACTAATCCCTGCATTATTTGGGTGCA
TGGGTGACAGCTCTGACAGGTTTGTGATGCCCCAGACAAATTCAGTAACTTTCAGTGAAG
CAAACCCATGAATAGATGTGATGGCAGCGTGTACACCTATATAATTCCAGAGCTAGTGAT
TATGTAACTTTTATATACGTCAGACCGAAGGAAGACAGAGAATGGAGGAACTGGGTGTTC
TTTCAGTAAAGAGCAACTGAATGAGACAGTACATCTTTTGAACTGGGGATATACTACAAG
GCAATGAGGGAGGCTGGCTATGAAAGTATTGAAAAATATGTTTGATTGCTGGGTGATGTT
TAGAGGCCCTAAGGTAATAGAAAGGAGACAAAATTGAGAGTCTGGAACTTATATGTACTT
TATTACAGTACTCTCATTTTCACCAAAGAAGGCAACCCATGTGGTGAAAAGACCACAAGC
ATTGGAGCTAAAGCCAAGTTATAGCTGTAGTTTTATATTTTGTGAGCCCATATGTCCTCA
GACAAGTTTCGTGAGTTAATTTCCTTTGTCTCAGCTTCCTCTTTTATAAAATGTGGGTGA
TATCATTGTCCCTTAAGATTGTTGTCAGCATTAAACAACATAAGTATATGAACCATCTAG
CTCGGTATTTGGCAATGGTAGGAGCTGAATAATTGTTAGCTCTTACCTTAAAAAATTATT

FIG. 1AC

TGTTAAAAGTTCCAAATGCAGCGTTCAGGAGAAGATATGGGTCAAGGTCATGGATGAGGC
AAACACTACAATTCAATAAAAATTGTTAGTTCTTAATTTATCTTAACTCAGCAACCGTTT
CTTGAGACTCTACTACATATTGAGTACTGAGGGAATAGAAAGATGAATCAAAGACCATTT
TAAAACATCTGGCATTTGCAATTCAAAATCAAGTAAAAATAAATACAGCCTTATGATTTA
TTGAGAAATGTCATGCAAGGTAAATGAACTGATTTTAAGCATGTACTTAGCATTCACACA
GATTGACAGATTCAGTGAAAACACGGCACAGCCTTCAATTATTTTTCTTTTTAAATACAT
ATTTGTGGACTTTATAGAAATACTGACAGTGTTTCCTCACCAATACCTATTTTCTTTGTT
GAGTGACTATTCCTTTTTCTTTTCAAATTAGTTTGTGTGGCAGTGTGGAAGAACCACCAC
ATGAGGACGGTAACCAACTACTTCATAGTCAATCTTTCTCTGGCTGATGTGCTCGTGACC
ATCACCTGCCTTCCAGCCACACTGGTCGTGGATATCACTGAGACCTGGTTTTTTGGACAG
TCCCTTTGCAAAGTGATTCCTTATCTACAGGTAATTGTTTTTAATGCTTTTTTGAAGCTA
CTAAAAAGAATGTTCAGCCATAGCGATGGCCCTTATGGTAAATTAACTAGTGAGTTGAGA
AATATATTTGCCTAAGGCATTGACAAACTGAAGGAAAATAATACTTGAGAATTTCTGGAG
AAATAAGTTAAGTTCTGGGTAAAAATTAAGCAATGAACTGCCAAATCATCATTAGATGCT
GCACAAACATTTTTGCACAACTTTTTTGATTACTAATTTGATTCCAAAAGTTTGATTTTG
CACAAACTTTTTTTATTCCAAATTTGATCCCAAAAGTTTGATTTTGCGCAAACTTTTTTG
ATTCCTAATTTCCCCATTGTTAAATAAGAAACTTGAACCAATTAATGATTTAACCAATTA
ATGATCTCCCCAAACCAATTATTGATCTTTCTCTTGAACCAATTAATGATCTGCCAGTCC
AAGTCATTGAGCATATTTGTTTTTACAAGTGATTTTATTTTATACTGAAGAATTAAGACC
TACTTGGTCAAATCAGTGCCATGAACAGGTTTTAGTGTAGATTCTAATTCAAACTACCGG
ATTTGGAATCTCCGTTCTGCCATTCACCAATTGTATGCTATCAAGCCAAATAGTTGTAAT
TCACTTATTTAAAAGAATAATTTAAATGAGATCTACCTCATATGGTTGCTGTGACCATTT
ACTTACATAATTCATATAAATAAGTTGGCACAGTGATTACCCTCTGGAAGAGATGATCTT
ATAAAAACAGTATATTCTCAATAAACATCAATTATCAGCATCAGAATCATCATTACTAGG
TGTTTTTCTTTCCTTAAGAGTGAAAACAGCTTCTTTTTCTATTTAATTGCCATTTCAGTA
ATTAAGAATGAATACTTTCAGAGATTAGTGTTCTGATTGTTATTATAGCTCTAAAATTTT
TGAAACAAAAGATTCATCAGATAATGTTCACATTCACTCATCCATCCTAAAAGATGGATT
TCCCTTAGGAATTGGACAGCAAATGAAATGGTGACCACTCTCTGCTTGTCTTCCCATAGC
TTTCCTGCACCCTCAGTTTTTACGCCATGCAGTCTCCCAGATGGTGCCTATAATATTTTA
AGAAAACAGAAAATAAGCTCCCAGTAACAAAAAAATTAGGGAGGGGTCACAAATAGCCTAT
TACTAGACATTATGCCGATTAGGCTTTTGGAATGAAATGTTGCAAAGAGATATTTAGTTC
AATAGTTCCTCAATTACCTCTTATAAAAAGAAGTGAAAAATTTTTAAGGTTAAACATTGT
TTATAGAATAGTAAGTGGAAAATACTATAGAAGTTATAAGCTCCATGCATATATTATGTT
TAATTATAAAGCTAGTTTGGATCAGCCTGCTGAAAATCATGAATGGATTACAAAACGAAC
AGTAGCACATTTTTTTGTGTGTGAGGAAAAACTACATGGGACAATAGAGAAAAATATTCT
CATAGAGGAAAAGTTAGTAAGAAATGAATGGCTCTGGTGGTGTTTGCATAGAGGCACTAG
GAAAGTAATACATTTCAGATAATTCTAATATTTCATTATCTCTGTGGTACTTCCAGAAAG
CCTTTTACCTCTCTTGGTTTCAATAACTACCCAGGAGAATATTTTGAGGATTCTCTTAAG
TTTTGGGATGGCTGCAGTTGCCCAGAATCTTCAACTGACTGGTAACATTTCATGTTCTCT
CTGTGAAACAGAAGATTCCCTGGTGGGAAGTGAAGTGATAAGGGCAGGTGCAGTCATGTG
CTAATGCACAGCGATAGCTTTCTGCAGAGCAGGCATCTCAGAGTTTCCTGTGAGTATTTG
CATTAGAGGACAGAATGGAAGCAGTGTAACCAGTGAGTGATGCAGAGCATGGGTATCTCT
TATAATCACTTACAGTCCTCTTTCACACAGCAGAACTATTTAACAAGTCCTACAGTTCAA
GGAATATCCTCATCTCTGGAAGGATTCTGTCTGCCTCTCTGCACACAGTGTCCAATCTAA
TCAATTCCTTAGCTGCTCCTCTTCTCCATAGAGCAAGGGAAAAAACTACTGGGTAACCAC
ATGATGCAAAAGACTAGATCCATTTGTTACCCCATCTAACATTACTTCTTGATGGAAAGG
TGTAAATGCACCAAGAGATTGGTGCACAGGTAAAACTAGTATCTCCAAATTCTTCATATT
TATTGCCTCATTTTTCATAGAATGTTCCCAAATGCAATGAACAGTGCCAATGGGCAATAA
ACATATAATTTAAATTTGAGCAGATTTTCTCCCTAGTTGTGACATTCTGTAACTAATGAC
TTATATCCCTGATATGATATTTATGTCTTACTGAATATTTAAAAACATGTTACATCATGC
CCAGCCACATTTTAAAGTTATTTGGTTGCATTTTAGATTACTTGGACGTTTATTAATTTG
CTATAATTTATATGTTCTTTTTCTTCTAAATACAATACAGCCTTTAGATTTATGAGTGAT
ATGCTGTAACGCATTGGCAAATGCACAAAAATCTCAAAAGTCTCACAAATGTTATAAAGC
TTAGCTGAATAATTAAAATGACTCTTTTGTATCTTTAATAATTGCATAACTCCAAGACCA
TTAACATGTATTCAGCTATTTGCTGAACAATTATCATGTATTTCACTTCTCTTCCAACAA
TGACAAGAGCATTGGTTACTTTTTCAGAGTGATTTTTTTAACTGCAGAAGACGCCCCTAC
ACAGAAAATGCCAGAAAAAAAAGAAGCCAAGTGAGATGTGGGAGGTGGGCAGTGGGTGGT

FIG. 1AD

CAAACAAGCTCCCTCTCTTTCAGTCATACTTTGAAACCTTTCTACCTATTAGTGCTTATC
ATCCAAATCTGTGATTTGGCAAAATTTTCATTTCTCCTTATAGTGAATCTTTAAGATACC
TTTGCCGTATCTATTTGCTAGTATAAAACAGTGGACTTCTCTACTAAAGGAAATCCCCAA
ACATTATCCTGTGCGAAGGGTGCCCATAGTATAGGTCAAAGACCAAGTACCTGAAGGCAG
AAGAAAGTTCCCATTATCTCACTCCACTTCATTCTCAACATTCATAATCCACACTAGATT
CATTTCTCAAATGACTTACTATTCAACAAACTTGAGCTAATATCAGAATCCAAATGAAAA
AGACACCCAGAAGTGCACTCTTAGAAGTTAAAAGCAACAACAAAACTTTCACTTATAATT
ACTTATGATAAAATGCAATTTTACATCACCTCCAAGAAAATCTTATACATTGCACATAAT
TGTATATTAATGTGTTAATTGCACAAGCAAATATAGTAGGTCAAACAATGAATATTAGCT
CACTGATTGTCAAGGGTTCATTCAATGGATTGGTTCATTCTACTGTTAGATACATCACAC
TAGCATATTCCTCCCTTTTCTGTGTGATGAAGGGCAGTGCTCCCTGGGTCACTATTGGCA
CTGGATGTCAGTCTTCCAAGTGAACTGATATGAATTGATTATTATGACCTAATGGCATTA
GGAAACACTAGAAATGACATTGATATTTGAACCATGCTACATCTATCCCATTTATCCATG
TTGATTAAATTAATGGATTATAAATTACTAAGGCTTGATGAACACTTTGTACTTCTAATT
GCTAGAGAGGATTGATATATCTCTAGCCCAGAAGCTATGAAAAGGCGACTGTGCGAATCT
ATACAACCATAGTTCTATTCCCAGGTTAGCAATGGTATTGAGGGGCCCTAGGTGCTTAAC
TTATTTGCAGAGAAGGAATGGAGGTTGTAGAGAATAAGGTGATACTGGTTTGAGAAAGAG
AGTTGAAGGTACCCTCAGGTAGCACTAAGAAATTTCTAGGAGTCACTAATCAACTTAAGC
CCATTCTCATAGAGTCCAGCCCCTTAAAATTACACTTAAAATGAAATTAGCCTCCAATAA
TTTAGCAAAGGTTAGGCTTTCACTTGTAATTTCTATGAATATTCTTCTCTGAAAAGCAAT
CTGTTCCAATTAAAATATAGAACTTCAGACTCAAGAATGAAAGATAAAACTAATAGTATC
ATCATCATTATTATTATTATATAATCATAAGAAATAGTAAACACACAGCACTTATATGCCAG
CCCTGGAATAGACATTTTCATCTCAACTAACTGTCCATACAATTCCATGGTTAGGTACTA
TTAATCATCCACATTTTACAGATGAGAAAACTGAGGAATGGAGAGGTTAAATAATCTCCT
TAAGATCACTCCATATGTCAGATGGGATTCATGCCCAGAAAACCTGGTTGCAGACTCGAT
TCCAGCTATACTCTTCTGCCTCTCCCATAGAGAAACAAAAGAATCATACTTGATAAGAAT
CTTATCCTGTTGATTTACTTCATTTAGCACACACACACACACACACACACACACGCAACA
CACAACACACAACACACACACATTAGGCCTAAAGCTGTAAAGTGAGTGACTCAATAGTGTGC
AGCTAGCTGATCAGAGAGAGAGAACAGATAGTTCATCCTGACAGCCCAGAGACTTTCTGC
ACTGTTGCACTGGATCTTAGATCTCTTTCACTCATTTGTACCTATAATCAACATATCAAC
AAGAAAGGTCCTCATGTAAAAGACAGAGATAACTACCCTTTCCACATATTATGAGATCAA
TATAACCAGGACAGAAAAATAGAAGAAGATGACTGGACTATATCTACTGCCTTCAATTAA
GGCTCACCACTATTAATGGATTAACAAATATTTGTTTTAAAGACACATGCAAGTATACGT
TCACTGCAGCACTATTCACAATAACAAAAACGTGGAATCAACCTAAATGCCCATCAATGA
TAGACTGGATAAAGATAATGTGGTACATATACACCATGGAATACTATGCAACCATAAAAA
AGAATGAGATCATGTCCTTTGCAGCAACATGAAAGGTGCTGGAGGCCATTATCCTTAGCA
AACAAATGCAGGAACAGAAAAGCAAATACTACATGTTCTCATTTATAAATGGGAGCTAAA
TGATGAGAACACATGGACACATAAAGGGGAACAACACGCACTGGGGCCTTTCAGAGGGTA
GAGGGTGGGAGAAGGGAGAGGATCAGGAAAAATAACCAGTGGATACTTGGATTAATACCT
GGGTGATGAAATAATCTGTACAGCAAACACCCATGACAGACATTTATCTATATAACAAAC
CTGCTCATGTACCCCTGAAATTAAAATAGAAGTTAAAAACAAAATATTTCTTAAATGCAT
AATGGATATCAAATGTTGTATCAGATATTGGGGACACAGTTGTGAAAAAAACAGAAGCAG
TCCCTCCTACCACAGAGCTTTGTTCCAATAGAGAAAACAGATGATAAATAAGCAAATTAA
GCAAATAATTTACTACATTATACATGCTGAAAGAAAAATAAATAACAATCTGTAAAAAAA
AATGTAAAAGAAATCAGAAGTCTTTTTAAAGGGAGAGGGGATTCTGAGAGTGATATCAGA
ATCAATATTTCATCCAGTATAAGAGAGCACATTGAACATAATTACATTAACTAATAATGT
GGATATATGAATTTTTAAAATTTTTTGTTGTTGTTATTTCCTTAAAGTGTCAAGTTAAAG
AATGATTTGTGGCATTGTTAATTATATACAAATTTTGACTGGGTGAACTTACCTAGTTTT
TGGAATCACATTGACTAGGCTAGCAGTGAGCAAACTGTCATAAGGAGATTCGCATACAAA
ATTCTCTTTTAATATGACTCGTAACTTTCCTTGGGTGCTACATGTTGAAAATGCACTGAT
GTACAAATAGCCCTTATTATTTGAAAATATGAAATAAGCTACCCATAATTTAAAAATGTT
AATTAAATATAATTTCAATCAAATTTCTATGTGGTAATTTAGAAGAAAGACATATTATTC
TTTATAATTGAGGCTTTTCCAGTTTGGACTAAACATATGTGTTTTTTTTTTTCTATATGA
GGGTATGATTTCTTCCAATCAATGGAAAAATTACAGGACAAAATAATTACAGTAATTATT
TAAAGAATGCCATATTATAAATTAAGACATTTGGAGTAAAAAAAGATTGCAAAGTTTTCA
TCATACCTTTTCATGTTTAACAATAAATTTACATTTAAAAGTATATTTCTAATATTTCAT
TTTTGTGATATAATTTCTTTTTAAATAGAAAGCACTTGCATGGATTGTTTATTTTTGGCA

FIG. 1AE

GCTTTGAATTTGCTTATATGTTGTGACTACCTTTCTCATATAGTAAATATATTAAGAGTA
ATTCTTTTAACAGCTGGTGCTTCTCTATTACTATGATCTTTCTTTTCTCTAGACCGTGTC
GGTGTCTGTGTCTGTCCTCACACTGAGCTGTATCGCCTTGGATCGGTGGTATGCAATCTG
TCACCCTTTGATGTTTAAGAGCACAGCAAAGCGGGCCCGTAACAGCATTGTCATCATCTG
GATTGTCTCCTGCATTATAATGATTCCTCAGGCCATCGTCATGGAGTGCAGCACCGTGTT
CCCAGGCTTAGCCAATAAAACCACCCTCTTTACGGTGTGTGATGAGCGCTGGGGTGGTAA
GTACCTTATGGCCCATCAACTGACATTTATATTACAGCAGCAAATTGAAAATTGGATTAG
CATAGCCATTGTAAAGCTGGGCTTATATATTTTATTGACATTTGTGAATACAGTTTTGCA
AGAGCATGAAAACCAACTTGAATTTCAAAACAATTTCACAGAATAACTCTACCTATCTGA
ATCCTTTGGAAATGTTATCTATTATTTTCTCATTTTCATATCTTTTGGATAGGAAATGAA
AGGAGATTATTCTACAATTCAGATTTGATTATTTTAGTTTTTCTTAAACTCTTTAAACAA
AAAGCAATATGGAATACAAATCCGATTATGTATTCTGGAATGATCCACGATTTATAAGAT
GGTTCAACACTGTGTTGTCTAGTGTCAGGGTCCCTAATGGGCTTCAAATACAACTGAATT
TTTTCATTTTAAGACCATGTCCTGGATCACATGGTCCTGGGAACATGGCCAGAGTCAGCA
TGTGGTTCTCTAAGTCAAATAATCCAAATTTGTTTTCTCTATTCATAATACATTATTGCT
ACTCGCATAATTATTATCCAGTTTAAGAATTATATTAATTATGAATCAATCTGGTTTCCC
ATCTGACAAGTATGATGTGAAATTTAAGCAATCAGGTTTGAAGGCTTTATGTTTCTTTGG
TTAGAAATTCTTAGAGTCAGTCTGAGGTTTTTGTGTAACAGTGAGAATACTGCTATCAAC
ACCTGGTGCTAGCACAAATCTGGGCACAGGAAAGAATGACAGAAAATAAAATAACCCTGC
ATTTCAGCATAGCATGCACTGATTCCAATATATCATATGAAATATATATTTAAAAAAAAA
CCAATCTGACCTCTTCTAGGTAAGTATACTAAAAATGGCTGATATTTAGAGAATTCATAT
GTTAACATTGTTTTTTATTAGAAAGATGTATCAAAACAAGCAGTGCACACCAGGGACTGA
TTAAGGATAATATTCTTAAATATTGTAATCTTTGAATTTCTGTTATTTCCTACCTTGGTG
TTTGTACTAGAACACCGAAAGGAAAAAAAGCCAATCACTGATATATTAGGCATATACTAC
AGGATATATCTACAGCAAGATAATATTTAAGAGAGGCTGGGATTATTTCATATATTGTTG
CAAGACCTATAATAACTAAAATTTTATAATTTGCTTTATCTATTACCCCAAATATCAAAT
ATCTGTCTTTTATTGGGATTTACTTTTCCTTTTTAACATTCCAACTTTTTTTTGCTGTAT
TTTTCTCTGTATCATTTTCAGTTTTTTCCAATTTTCCAAATTAATAGTGCAGACAAAAAA
AAAATCAATGGAAATTTCCAAAATGGTAGGAATATTTATGAAGTGTCTTATGTCCCATTC
ATTTAATGCTCAAACACCACCTTGAGAACTTAGTATATGTCAGGCATTGTGCCCACCTGG
AGAGAAACAGACTCTGCTTACGGGAGCACACTCTATATAATAAGGCTCAAAGGCCAATAA
ACAAATTTTTATAGGGTAATCAGTATTTTAATATATTTATATACAAAATGCTGAGAACAC
AAATGAGAGAACAAACTCAGTTCTGGCCATTTGAACAAAAGTTTACAGAGGAACTGCTAA
CATTCCAGCAGAACATTAAAGATAAGCAAAAATTCTCCAGACTGAGAAGAGGGAAAAGGA
TGTCCAGAAAGCAAGAAAATCCACATCATGGATACTACATTACAAAGCAGAAAGAGTGAA
TCAGCACTTGTAGTTTCTGGAACATAGGGGCAGGTAGTGTAAATAATTGAATTTTGAAAC
AAGATGGGTTGGGGACTGACTGTGACATGTCTCTTATACGATCCTTTTACACTGGTTTAT
ATTTAGAAAGCCTAAAAAGGTCTTTCTCAGAATCCTGTATTAAACTCGAGACTAAATTTA
ACCCTAGAAAGATTATATTATTTTTTCAAGATTATGAAGCAAATAGGTACATTTAAATCT
AAAGCTTCCAACTTGTAAGTTGGGATTCCTTAAGTTTTATAGGGATTGCTATTAGATAAA
ATATAAAAAATATTTTTCAATATGTGTCAGCAGTATTTTCTCTAATATTCCGGCAATTAGT
TTCACTTATATGTTTATGGGTTGCTTTTATAAGCTTTTCTTTTTTTAATGTTTCCCTGAA
TAATCAAGTAACAGTAACCTCCATTAACAAAAAGATTGCAAAGTCATGGATTCCTGTTCA
GTTATTATGATTATGTAAATAGACGTATGATTTTTAAATTACCTCTGAGTGGTAAATATA
AATACATAAAGCTCATTTCTACTCTGATATTTTATTACATAACTCTAGCATGGACATTTT
CATTAAAAAAAGGAAACAATTGTTGAATATGTAAAAACCTAAACTTAGCCTTCAGAAGTC
ATTTAAGAAAACTATTTGAAGGTGATTTTATAATAGCCTATAATTAAATGCTTGTAAAGA
CTAAAATTAAGTATTATTGGACTGAATTGATTAGCTACAAAATCCAACTTAGTAAAAGCT
ATACAGTCATTTAAATATTAAATGAAATTGCTAAGAATATTTTTAAGAAAAAATAATTCA
AGGCAGATTTTTATCTTTCTTATTAGATATTTATTATGATGATTTCTACATAGCATGTAA
AATCATTGTTCATGTAAACTATTTATAAGTCCATGTTCGACTTATAATGTTAAACCTTTG
TATATGTGTGATTGTCACAACTTTTTAAAAAACCATAGGAAAGTATATTTTACAGTGTCA
TCTCTCTAAATTCAAATATTTTTAAAGGCCAACTGTCATTTAGCCTGATTTTTAAAACTA
TTGTAAAATATCTTCTATTTGAGATTAATTCATAATCTGTGTTTCTTATCTTTATTCTAA
GTTAAATCAATAATGTAGTTATAAAAGTAGAGAGTAGAATCATAATTATCCTACAACCAA
TGTGGCAGTGGAAAAAAATTTGGAAAAGCAATTTGGTCAGTTGATACATATCTATCAAAT
AACTTTTGGAAAAGTTCTGTAAATGCTGTTTTACTCATGGTGCAAAATAACTGAGAACTC

FIG. 1AF

TGTCTAACTAAAAAATTTACCAGCAATATGTAATTATATATGGATAAATGATTTCTAAAA
CTAATTATATTCATTATTGCCTATTACTTCTTCATAAAAAGAACCATAAGCCATGATTTC
TGGCAGACACACACAACACTCAAGAACATATAAATAATGTAAATACTTATTTTAATAACC
TTTAAAATATACATTTGTATGTGTTCACTGTTTGCTTCAGTCACATCATTTCATACTTCT
AAAATTATTAAATTAACCCACAATTTCTTGCTTGCTTGGTTTGTAAATGCATAATTCTAC
AGGAAAGATCCTACAGAAAGAAATTCTTTGCTGGGTGTGGTGGCTCAAGCCTGTAATCCC
AGCACTTTGGGAGGCCGAGATGGGCGGATCATGAGGTCAGGAGTTGGAGACCAGCCTGGC
CAACATGGTGAAACCCCGTCTCTACTGAAAACACAAAAATTAGCTGGGCATGGTGGTGGG
CGCCTGTAATTCCAGCTACTCGGGAAGCTGAGGCAGGACAATCGCTTGAAACCGAAAGGC
GGAGGTTGCAGTGAGCCGAGATCATGGCACTGCACTCCAGCCTGGGCAAAAGAGCAAGAC
ACCATCTTCAAAAAGAGAGAAAATAACTCTTTTTGTACACTCAATCAAAGTTATATTTTC
TTCACTATTCATTCATCCAGTGTTTAATTAGCATGTACCCTTGGTCAATTGTTCTGGACA
CTGGAGATTAGTAGCATCTCTCTTTTTGAATATTACTGACAAATTGTTCTTTGGTAGGCT
AAAAAAAAAAAATGGAACCATTTTTACAGTCAAAGTAATTATGGCATCTGGCCTATTATG
AGGTTTGAAAGCATATAAATATGTGTATAAGTCTATTAATGGGAAGATTTATTAAACATA
TTTATTAGGGAGAAGATAGTAAAACATATTAAAGATTCAGGTAAACTTAATGAACCCCTA
AACTTTGAAAAGACATTCCATGTTGAATATTGGGAAATTATATTTAATTTACTTGTTCAT
TCAATTCCTGATAAGTGTACCATGAAAGAGGAATGTTTCTAGTTTCTAGATAATTAAGAT
AACATGCTGGCTGAATAATGAACCTTAAGTCATCTGAGAGAAATTAAGTTTTGCCTGTCA
AATATACAATATAACTCTTTAATCTCTGATTTCAAAGACTAAAGATCCACATTTGTTCCT
TATTAGTTAGTTTCATATATATATATATATAAAATTTATTTAGATTGTGCTTATTCATCAGT
TGAGTAAAAACAGTAATTTTTAATGATTATCAATATTTAAAACTTTTTTAAATTAAAGTA
ATGCTTATGTGAAACAAATTTTGTGTAGTTATATTCTAGGTTATATACAAATGTCTTAAA
TACATTGAAGACATTGCTTATGAAGTACAGAAAGACTTCAAAGATATTTTCATCACACAT
AATTTAAAATTTCAATGGCATATCTGAGTTTTTAATCAGCTTAGACTATCATGTTTCCCT
AGTTATCTATTATAATCTCCTTATTCAAACAATCTATCCTACCCTGGAAGGATAATTTTG
CTTGATCTTTTTTCCATATCAGTGTTCATTATAATAATTTGCATTTAGCAGTCAATTACA
TATTTTTTCTAATTATTCATAAATATACCAACCACATAGGAGCTTTTGCTACCATCTATT
CAAAACGCCAAACTGTTATCACAGTGATGCTATCCATAGCTGCAGTGGAAAAAATTTACC
TCTCAAATCTACTTTCCTCTATCCACTCAATTGGTCTTATGCAGACAACAGGGCTTCGCA
GGTATGTAAGCTTCAAAGTTATATAGATTTTGTCATGAGGAAAGCTCATGTGACACCTCT
TCAAAACAAATAAAAGTTCAAAGCCTCTTAGGTGCCTGGGAAGTGCTGAGATCACTTTCA
GATTCCTTTGAAATTGGCCCGCCATATGCTGTGTAGGCTGTGGCACTTCAAAGGGAAAGA
CTGTTATTTCTCAAGTCAGAATGCTTGAATGTTATCACTTTTTATGTAACTGGCCTGCTT
TACAGGATCAACTTGAAAGAAAGTTGGAAACTGATGAGGTAGGTGAGTGCTACCTGGGCC
AGAGAGTAGCTAAAAATGACACCTCAAATTGGTCTCTTAGACCTGCCAACACATGCATCC
TACTGACCCTGCTGAAGACTGCAGCGGATAAAGACATCTAAACCAAAAGAGAAGATGGGT
TTAGAAGCATGAATATGGAGAAAATTAGACTCAAACTCAACTGCATCTGAAAGACAGCCT
ATGGAAATAAGATTGTGGAGGATATTAAACTCATAAATATGTTAAAATATATCCAGCAAG
AATCAAATGCATGATTGCTCAATAAATATTATCTATTATTATGACAATCATCATGCTTAT
TATTGATTAATCCTGACTGTAAACTGCTCTTATCACAAATCTGATCACATAACCAAGCTT
TCATGCTTCTACATCCCCTTTATGAAGTAATGAAAAGAATAAAATACATAGAGGTAATAG
CATTATTCCTCAACAATACTATGGGATAAACCCCCTTGTCAATAGAAAAGTCAAAACAAA
GTATGTAAATTTTAGAAGAAAAACAAAACAGCTCTGTTGTGTTAGCATTCAATTAGAATT
ATAATGAGTTAATTACATTTAATATCTATGGAATCTATGCAAGATATATTGCTTCCTCTT
TTACATTGCAGTAAAAGTAGGTAGACCATTGTGATATATTCGAATACAAGTACAAAAATA
TCTTCTAAAATCTACAGGGAACTCAAACAAATCAGGAAGAAAAAATGCAAACAATCTTAT
CAAAAGTTGGCTAAGAACATGAATAGACAATTCTCAAAAGAAGGTATACAAATAGCCAAC
AAGCATATGAAAAAATGTTCAGTATCACTAAGAATCAGGGAAATGCAAATCAAAACCACA
ATGCAATACCACTTTTTTTTTTATTTTTTATTTTATTTTTTGATGGAGCCTCGCACTGTC
GCCCAGGCTGGAGTGCAGTGGCATGATCTCAGCTCTCTGCGACCTCCACCTCCCAAGTTC
AAACGATACTCCTGCCTTGGTCTCCCAAAGTACTGGGATTACAGACGTGAGCCTGTAATT
GGTGTCTGGCCAATACCACCTTACTCTTACAAGGATGGCCATAATCAAAAAGCCAAAAAA
TTAAGGACATTGGAATGAATGTGGTGGAGAGGGAACACTTTTACACTGCTGGTGGGAATG
TAAGCTAGTACGACAACTATGGAAAACAGTGTGGAGATTCCTTAAAGAACTAAAAGTAGA
TCTACTATTTGATCCATCAATCTCCCTACTCTGGTAGCTACCCAGAGGAAAATAAGCCAT
TATACTAAAAAGATACCTGCACATGCATGTTTACAGCAGCACAATTCGCAAATGGAAAAA

FIG. 1AG

TATAGAACCAGCCCAAATGCCCATCAATCAATGAGGGAATAAAAATATGTGGTATGTATA
TAGCATAGAATACTACTTAGCCATAAAAAGGAACGAAATAATGGCATTCCCAGCAACCTG
GAGGGATTTGGAGACCATTATTCTAAGTGAAGTAATTCAGGAATGGAAAACCAAACAACA
TATGTTCTCACTCATAAGTGGGAGGATGCAAAGGCATAAGAATGATAAAATGGACTTCAG
GTACTCAGGGGAAAGTGAGGGAGAGGGGGTGAAGGATAAAAGACCACAGATTGGGTAAAG
TGTACACTGCATGGGTGATAGATGCACCAAAATCTCAGAAATCACCACTGAAGATTCATG
TAACCAAACACCAACTGTTTCCCCAAAACCTATCGGAATAAAAAATTAAAAAAATACATA
CATACAAAAATTCAGATTCCCGACATAATATATAAATATATATTATATGTTATATATAAT
ATTATATATAAATATATAATGTATTATAGTTATATATAATATTATATATAAATATATAAT
GTATTATATGTTATATATAATATTACATATAAATATCTATTAATATATATTATTTATATT
ATATCTAAATATATAATATATAACTTATTATTATATATTATAATATAACTTATATATTAT
ATATAATATATATAATATAATATAACTTATTATATATTATATATTTATATATAAATAAAA
TATGTACTATATTAATATATGAATATATCTAATATTAATATACAATATATAAATTTATAA
ATATATAATGATTATATATTATATAATATATAAAATATATATTATGTAGGGAATCTGAAT
TTATTTATGTATTTATGTACATATATAAGGTAGGGAATATATATATATGTATTAGGTAGG
GAATATATATATATATATATATATCTTCTAGAGCATTTACAAAGTTAGTAATCAATATAA
TTTAGAAAAGCTAAAATATTAAACCACAATGCCATGAAGTGATTAATCGACTTATTCGTA
AGTGTCTAATCTGTGATGTGTATCATTTGTGTACATAGGATTAATTATAAATAAAAAATT
ACTACAGTCCTAGAGGTGTTTATGCTTAATAAGTGAGAAAATATTCATATTGGATTGGAG
AAAATAAATGTTATAAAGCCTTAAAATTCTCATTTTTATTAAAAGTATATACATGTATTT
TTAATAAAAGCATACACACACCACAGACATACTATGCTTAAAGAGGAATTTTGTATATGT
TCCAATAAGTCAACAAAAATAATCATTGTCAAATTTGTATTGTATTTAGTTTTCAAAATT
TTTTTCACATTTGTATTTGGAGATACAACTGAGAATAGCCTCCCATTTCTCAGGGAACTT
ACATTCTAATAAGGAACAACCAACTGAGTTTATATTTCTTCCCATTTTAACCAAAGCAT
TAGTTTTTAGGTTTTCATTGATTCATGTCCCTTTTTGTAAATAAAAGTTTAGAACAACCC
AAATTAATTTTGTTAATTAGCCAGATGTAATCAAGTCAAATAAAGGGCCTTTTAATAACT
GAACACTTGACTTTGGGTAGCACAAATTAAGAAATAGCTAATGCTTATTTTTCTGAGTAC
ATTAAGTGAAATTACGACTTCACATTTGGCATGTGTATACCCATATACTGAGTAAAATAA
GTTGTTAAATATTATGAATTATTTTTCCCCTTTGCATACATAATATGACAATGAAATCAT
ATAAAAGGTAAATATGCACTTTGAAGAAAAGCATTGACATGTATCTTTTTTAAAAGTCCA
TCAATTGTAACGTAAGGTTTGTTGTTTTGACTTTCATCCTAGGTGAAATTTATCCCAAG
ATGTACCACATCTGTTTCTTTCTGGTGACATACATGGCACCACTGTGTCTCATGGTGTTG
GCTTATCTGCAAATATTTCGCAAACTCTGGTGTCGACAGGTATATAGTTTCAAATATTTT
GCGTGCATTATTCCTCCACACATAATTTGTTATTTGTTATTCCTTCCAAATATTTTGTCT
GTGCTTTTTTTTTAGGATGCACTTATAAACAAAATTTAAGAATGCATTGAACCAATATAA
CATGTTCATAAAAGTATTATATTGTGTGTTCTTTTAAAGTAATGAGAACCCAGACATAGA
AATATGTCTAGGCATTTTTAGAGTAATATTCAGGAAATGTATTTTATAAACTGATTAAGT
ACTTTACATTTTAAATAAAATTTAACATCTGTGATTAATTGTCTTTTGTCTAGGAATAAC
ACTAATTTCGCTTTCTATGAGAAATAGCAAATAAAAATTCCTTTAGAGATTTTTGAGACT
CTAAGTCTGAAAGGTTATATTTGTAATCAGATTTATTTAAAACATTGGAACATATAGGTT
AAATCTCCAACTTCAAAGATCTTATTTTTTAGAATATTATAAGAATCAGGCAGAATGTAT
AATTTTAAAAACTGTATATAATGCTGATTTGGGGTTACTACACTTTGTTAGATAATTCTG
CTGTATCAGTGAATGTTTGTATTCATTCACTCAGTTATTCATTCCTGAAATACATATCAT
GAACTTTTCACATACATGTCTCACACAAAAGCTAAAAATTCTACTTTTTGCCATTGAGGA
ATTCATAGTCTAGAGGAGGGGCATCATCAGATGCAGGGCGAAAATTACTTTAAATATAAG
CACAGAGAATCAGAGCAAAATGTACTAAAACCATATCTAATACAGGAAAGGTAACATTTA
ACTTAAACCTTGATGATTTGAAGGATATTACCAACAAACACATTTAGTGGTTTGTAAGAT
AGAGACAAAAAGGATATGGCTCAGTCTCTCCCATTTTGTAAAATGTATCTTAAAATGCCA
CAATTCTTAGAGATGTATTTCTCTCGTTCTCTAAACTTACTGCCGATACTTACTTTATCA
GGCTTGTGGAAGGACATGCCATTAGTCTGTTTTCTCTGACACATTTTATCCAACTGAAAA
GATTTACTGGAGTCACCTTAATTCATTAAAAAGATTTCACAAACACTTTATTTGGTCTTT
GAGGATGTGTCTTTGTTTTTTTAATCAACACTTGTTATTCAAAGCATTTTTCAAGATCAT
CTTTCACTGACTGGATATGAGCAACACTCATTTTTTTTAACACTATATGGCTCATAATTT
CAATATTTTCTCTTTTCCTCTGCTATTACAAAGAAGTCATTTCTTTTATGACCTTACAAG
TGAAACCAGTAGCAACATTTATTAACATTTTGTTTCCCATCATTTTTTACTATAAAAACT
AATGTGGACCACTATAAAATATGAGTGGTGATTTTCTAGATGTTGGTGACAGTTTTCTCA
GCACTCTCCACCTCCCTATGAAGCCAATGCTTATATTTTAGGGTGTTTGTTACTGCAGCA

FIG. 1AH

TCCTGCTTCCTAGTACCTATTATTGTATCTGTCAGGTTTTGCTAGGTTATTATTCTTCTA
TTAAAAAATGTGGTTTGCAACAACAGTTCTGTTTCACTCCTATTACAGGTCAGTGGGGAG
GGCTGGCTGGGGCACTGTGCTCCATTTGTTTTCTCATTCCAGAACCTAGTCTGAAGAAAT
GGCACTTTCTGGGACATGGCATTCTGAGACTGAGAGAAAAAGAAAACTGGAAGAAAAGTA
TATTTTCTTTTAATGTCTTTTATGAACCGGCATGTGTTACATCTCACTTTTCATTGGCTA
AAACAAGTCACGTGGTTAAACTTGATCATGAAGAGGGGACACATTCTTCTCTGACAGAAA
GACATCACACATCACAGGGTAATGGGGAGCTTCCTACAAGCTGGGGATGAATGATCTGGA
ATGATACACTATACAGAAGTCAAAAACACAAGGGCCAGACTGCATGAATTTAAATCCTGA
CTCCACCAAGTAGTAGTGACATGAATTTTGTAAATGGCTTAAATTTTTGTGACTCCCTTT
ATTAACTTTAAAATGGGGTTGTATAGCATCTTCCTCATAGGTTTGGTACATGCATTCAGG
TGTGTCCAAGGGAGAGAACACCGTCGTGGGTTCTCAGTTTCTATTTCTATTTGGGCCAGT
AAAACCCCTTCCTATCCCTCTTTTCTGCTTATTACTAGAGACAGAAACTAAAAACCAGGG
CTTCAGGCTGCTAAAAGCCTAAAACAAAACAAAACAAAACTACAACAACAAAATAAGGTG
GGTTGGACAAGCTTGCTTAGATGAATTAACTCAAGTGCCTAAATATAGACAGTGCTCATT
AAACAAAATATCTTAATGGATGTTGTTTAATAATGGCCTCTCAACTAATTGTACTTACAT
TTAAATAGCAAGCATGTGTTGAATTGGTATATGTGACTATTTTTTAAAAAATGCACATTG
AAATACCAGTATGGTGCTTCTTATTTGTCTGGTTCTTCTACTCTACTAAGATAAAGATAG
TCTCGCTGTCATCTTTGTATCCCTATAAATAGCACGTGCTCAGCACACATCAGTTGCTTT
TTTCATAAGAACAAAGTGAGTAGAATAGGAGAAAGTGCTGGGAAAGTTTAGAGAGGACAT
AGAGAATCTATTGCCCAGTTACTCCGATAAACATTTGTAGAAATGGATTAGAATCTGAAA
AATTTCTTGAAGGGGAAAAAGCAATTAATGAGCATGTAGGAATAAAGATATTTTAGATTT
AGATTCAGATTTTGTTGGGGAATGTTCAGTGTTAAGATTATCCCCTATTTCCTTATTTTT
ACTAGTTAGTGTGCATTGTATAAAAGGTATGCTTATAATTTCTTATTCATTTATTTACAA
ATTGACATACCTTTAAAACTCTTTCAAGGTTGCAATGTATCTGTCTTGTTACTTTTACAT
GGTAAAACTTTACCATGATACCATGGTTACCCTAAAGTTTACATGGTACCATCAGAGAAA
ATGTTTTAAAAAGTTTGTTAAATGAATGAGTGACACCAAAATCCAAACATTTTAATTTTC
CACCATTTAAGCATATAGTTTGACATTTCCCAAACTCTAAAATAAATTTTAAAATAAATT
GCATCACAGATTCATAAATAATCCACATTCTTTTCATGAATTATCCTCATTAGTACAAGC
CACATGATTCAGAAGATTTGCAGTAAAATGCTTGGGCTGTGAAACTAAAGTCATTTACAA
AACAGATTGGAATGGAAAATACCAAGTTCAGCTGAACTCACTTTAGCAGCCACAATAAAG
TGAATTAACCCCAAATGCGTGATTACATAGAATTCTGCTTGAGCAACTCTCAATTTCCAA
CTGTTAGTGTCTATAAACAAAGTTGTAAGGCATTATGCGTGCCATAGGCTACATCAAGTG
AGCCATCAAATGAAGAGCTTGTCCTATTTGCTTAAAATTACAGAGATGCATGAAATCTGT
TATGTACTTTTGAATTAGTAAGTGTAAGATTATTAGTGAGCAAATTGTGTGTCCTTGTCT
GACTTTCTCAAGAAGTTTAAGCCTCATTAAAAGAATTAGCTAATGCATTGCTGTGAACTA
CTTAAATTCTCTCTCTCTCTGTTTTTTTTTTTTTTTTTGGCAATTCGACTCAGAGTACTC
AGGAAATTCTACAGATTATTTGCTAAAACTTATTTTTTTAAAGAACTTAGCTTGCTTGAC
TCTTTCATTTATCTGTCAGCATTTTTTCTAGTTCAGACCCTTCATATAATTCAACACTAA
ATCTTAATCGTCATGTGCTTGTGTTAATTTATTTCACATTTATTAAGCACGTACTCTGTG
TCAGCTATGGTGTGAGGTACTGAGGATGGACTGTAATAGATATTTGGGTCTGAAACTATA
GTTCAGCTTCTCAGGGCCTTTGAAAGACCTTCTTGTTCCCAGCTCTTATCACAAAGTTTT
CTGCTGCTCTTTATTCAGCACTCTTCTAAGGGAACTTAAGATAAATAATATTTGATGATG
ACAAATCAGTCTAGTGTGAGAAAATAGGCAGCAAACAAATTACAATTGCAGGGGCAGAAT
CAGGAAGGCAGTAACTCGAGTCCATACAAAAAAAAAATAAGGAGCACCAGTAAAGGTAACT
ACATAGGTAAATACTGTAGACAGAATAAACATATTTATCTTCTGTTATCTGATGTAAAGA
ACAACTGCATAAAATAATAGCTATAAAATTGTGAAGATTCACCTTATAATGTATACAGAT
GTAGTTCAAAAAAGGAGGAGGAAATGGAGCTGTATTGGAGCAAATTTGTTTTATACTATT
GAAATTACATTGGCATAATCTAAGCAGCTTGTTTAGATTAAGTTGCTAATTTTAATTCCT
GGTGTAACCACTAAGAAAATAATTTTTTGAAGAATGTAGAAATATAGGTAAAGTAACAAA
AGAATTAAAATAGTATACAGAAAATATTTAACACAAAATAAAGCAGTAGTGAGGAAATAG
AGGAAGACAAGAGATACAATATATATAAGTCACAAATAGTAAAATGGCAGATATATATTA
TATTTTCTTAATAATTAGATTAAATGTAAATGGATACAATACTTCAATCAAAGGGCATAG
ATTGACATAATAGATAAAAACCAACCAATAATTTAAAAAAACCCATGATCCAACTTTATG
CTGTCTACAAGAGACATACCTTGTATTCAGATATACAAATAGGTTAAATGTAAAATAACA
GAAGAAGTACTAAAATAATCACAAAAAGGGAGTTAATGTGGTTATACTAAAATTAGACAA
AATAAATTTTAACAAAATATTACTATACATAGAGAGGGACATTTCATAATGATGATGGAG
TTGATCCATCAGGAAGATATAAAGTTGTAAACATACATGCATTTAGCCACTGAAACCCA

FIG. 1AI

AAATATACCAAGCAAAAAGTAGTAGAATTAAGGAGACAAGTAGGCAGCTAGACAATTATA
GTTGAAAATGTTAATACTCACTTGCAGTAATGGATAGAAAACACAGGCAGGGTGCGGTGG
CTCACCCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGTGGATCACGAGGTCAGAAG
ATTGAGACCATCCTAGCTAACATGGTGAAACCTCGTCTCTACTAAAGATACAAAAAATTA
GCCGGGTGAGGTGGGGGGCACCTGTAGTCCTAGCTGCTCAGGAGGCTGAGGCAGGAGAAT
GGCGTGAACCCGGGGGGTGGAGCCTGCAGTGAACAGAGATCGTGCCACTGCACTCCAGCC
TGGGCAACAGCAAGACTCCGTCTCAAAAAAAAAAAAAAAAAAAGAAATAAAGAAAAAACAG
GCAGAATAGCAACAAGGAAATAAAAGATTTAAACAAACTATGAAACCACTGGGCTTAACA
GATATTTTAGAACACTCCACCAAAAACAGAAGAATGCATATTTATCCCATTTGCACATAA
AACATTTTCCAGGTTTTCTGACTAAAGTCAGAAACAAGACAAGTATGTCTGCTACAACCA
TTTTCATTCAATGTTGAACAGAACTGATTCTTTTCAGGGCAAACAGGCAAGAGATAATAT
TAACAATAATAAAAAATAAAAGGCATGACGATCACAAAATAAGAGGTAAACTATTTCTAC
TTGTAGGTTATGTGATATTTTATATAGAAAATCCTAACGAATTATTTTGCAAAAAAATAC
ATTAGAACAAATAAATGAGTTCAGCTAGTTTTCAGGATGAAAGATTAATATATATACAAA
AATCAATTTCATTTTTATACATTAGCAAATAAAAATTTAAAATGAAATTAACAAAAAATA
ATTTAAATAGCATCAAAATTAATCAAATACCTAGAAGTAGATTTAATAAAAGAAGCTTAA
TAAGAGACTTCATCCAGGCTTGATTGCTTATGCCTGTAATCTCAACACTTTTGGGAGACT
GAGGCGGGAGGATCACATGAGGCCAGGAGATCAAGACCAGCATAGTCAACGTGGTGAAAC
CATGTTTCTACTAAAAACACAAAAATGAGCCAGGCATGGTGGTGCAGTGCAAGACTATAA
TCCCAGCTACTCAGGAAGCTGAGGCATGAGAATCATTTGAGCCTCAGAGGTGGAGGCTGC
AGTGAGCTAAGACTGCACCACTGCACTCCAGCCTGTGTGGCAGAGTTAGACTCTTGTCAA
AACAAAAAAAAATTCTTCAGCATAAACATGTATATTTAGGGAATGTCCAGAAATTATAGAG
ACATGGATTCCATGCAGCAGTTATAATTCCCTAAAAAGAGAATTATGAATTCACTGTATT
GCTGAGGATTCTAACATAACCACCAAAGATCCAGGGAGAAAATTACCCTATTTTTGTATT
TAAAAAGATGCATTTATTAAATGATGTGGTACTAGTCTCTATATAGGCAACAAAAATAAT
GAAAAGGAAATAGCTCTGGATTATTAAAAATAAATAGTCTGTTAATCAAATCAATTAAAT
AGATAATGTTCCTTCAACATTTTCAAGTCCTATACATGAATATCATTTACAATCATAATT
ATTAGCAACTTCAATGAGTAGGCCACAGTTATACAAGTTTCTTGAGTCAGTTTGGAACTA
TTTCCATTCAAGCAACATATAGTCCATTTCTGTAACATTTTGTTCTCCATCATTATATTC
AGTCTCAGAAAGGTTACCAACACAGTCCTTGAATCACATGTAGTACAGGTTAAGCATCTC
TAATCCCAAAACCTAAAATTCTAGCTGCTCTAAAATCCCAAACTTTTGAGAGCTAACATG
ATGCCAGAAGTGGAAAAGTCCCCTGCTATCTCATGTGACAGGTCGTGTCAAAAGTCAACA
AAAACTTTGTTTCATGCCCAAAATTATTAAAAATGTTATATAAATTTGTTTAAAGACTAT
TTGTATTGGGTGTTTATAAAATGTAAGTAAGTTTTGGGTTTAGACTTAAGTCACATCTAC
AAGATATCTTTTTATGTATATGAAAATAATCCAAAATCCAAAAAAACTCACATCTGAAAC
ACTTTTGGTCTCAAGAATTTCAGATAAGGGATATTCAATCGGTACACAACATATACACCT
ACAATTACAAAATATCATTGAAAAAACTTAAAGAAGGACTACCTAAATTAAAAGATATTC
TGTGTTTATGGATTGGAAGATTCAATCTTGTTAAAATAGAAATAATCTTCAAATTAATCC
ATGAATTCAATACAATTCCTATGAAAATCCCAGATGGCTTTGTATTTTGGACACAAATTG
ACAAACTGATCCTAATATTAATGTGGAAATGCAAGGGATAGAGAAGAGCCAAAATAATCC
TGAAAAAGAAATGGAAAACTTACTTCCTAATGTCAAATCTTAACAAAAAGCCACAGTAAC
TAAGACGGTGTGGTACTTCCATACAGTTAGTCATATAGATCAGTGGAATAGAATTCATGG
TCCAGAAATAAACTCATATTTATGATTAATTGAGTATTGATAAAGGTTTTAACACAGTTC
AATGGCAAAAATCATAGTCCTGACAACAAATGGTGTTAAGACAATTGTATATCCACAAGC
AAAAGGATGGAGTTGAACCTCACCTCACACCACATTCAAAACTTAACTCAAAATGAATCA
TAGATTTATATGTAAGAGCTAATCTCTTAGAAGAAAACACAGAAGAAAATCATCATGACC
TTGGCTTAACCAATAGGTTCTAAATATAACACCAAAACCAAAAGCAACAAATGACAATGT
AGATACATTAGACATTATCAAAACAAAAACTTTTGTGCTTCAAACTGCACCATTAAAAAC
GTTAAAAGTCAGCCCATATAAATGCAGAAAATATTTGCAAATCATATATGTGTTAAGGAA
TTTGTATCCAGAATATACAAAGAACTCTTATGAATTAATAATTTAAAAAAATTACAAGTA
GGCAAAGACTTGAATAAACAATTCTGCAAAGAAGATATACAAATGGTCAATAAGCACATA
AGAAGGTGCTTAACATCATTACTCATTAGAGAAATATTAATCAAAATCATGAGATACCTA
TTCACACTCAACAGGATAGATTTGTTTTAAAAGGCTGTAATCATTATTGGTAAGGATGTG
GAGTAATTGGAATCCTTCTACATTGTTGGTGGGAATGCAAAACGATGTAACTGCTTTGGA
AAACAGTTTGGTAGTTCCTTAAAATCTTAGAGAATTACCACATTACCCACTAATTCAATC
TCTAGTTATAGACCCAGAGAACTGAAGACATGTTTACACAAAAACTTACACATGAATGTT
CATAACAGCATATAATTCATAGTAGCCAAAAAGTGGAAACAACCCAAATGTTATCAATGA

FIG. 1AJ

GTAAATGGAATAACTCATTGTTCTATATGCAAGCAATAAAATATTATTCAGCTACTAAAA
GAAATGAAGCACTGATATATGCCACAAGATTGATGAATCTTGAAAACATACTAAGTGAAA
GAAGCCAGGCACAGAAGGCCACATATTACATAATTCTATTTGCATGAAAATGTTGAGAAT
AGGCAAATATATAGAGCCAAAATAATTTGTCCTTGGCACGGGCTGGCAGAATGGGACAAT
GAGAAGTGACTGCTAATGGATTTGGAGCCTCATTTGGAGGTGATGAAAATGTTCTAGATT
AGTTAGTGATGATGTGATAGTTGCACAACTCTGTGAATATTCTAAAAATCATTTTTTGAA
CCCCTTAAAGCAGTGAGGTTTATGGTATGTGAATTATATCGCAATAAAATGTTTTCTTTT
AAAAAGAAAGAACAAAAATGATGGGATATTTTAAAATTTTAAAAATTGAAGACTTTTTTT
TTTTTTAGAAAGTTCTGCTGCTGAAACCACAGGGAAGCAAAAAAGGTTGAACACACAATT
TGACATGTTAATGTAATGAGAGACTATAATAGGAATTATCCACGGGTTGTTTTATCTGTA
CTTTCTGACTAAAGTTTTTTTCCGTACTTCTATAGACTTTAAAATGGTCCATAGATGTGC
AAAAAATGAGAGAACCTATTCCATGAAACCATATATCAAGTCCCAGAGAGCAGAGGGAAA
ACCTTTTTTTTTTTTTTTTTTTGCAAAGAAGAAGTCATAGACTGTGTGAAAGAATAATGT
TGCGAGACAACAGATCTGGAGTTGGACAGGGGCAGGAGGCATAGTGAGAAGATCAGTTAT
TGCAGTTGTCATCCATAAGGGCCATCTGTACACTCTGAAAGTGGAGCTATTCATAGTGAG
AATGATGTTAAGAAAAGGAACAAATAAAATTACAGTCCTCGTTATAAGAATTTAGCATGC
AAATCTTATCAGAGCAGTACTGAGGTAAACAAAAAGTGTCAAGAAATCATGGGATTTAAT
GTGAAAAACTCCCTCAGTGTTGGAATACAGTCATCTTCATATGGTGGTGGGTGTAAGGGG
CAAGGAAAATTTTCATGGTCCCTGCTGAACAGGGAAATGTAAGGGGATTATTGTTTCATA
GAAGACCGCCAGTGCCTACCAAATATCTGTTATACTCTATTATGATGAAATGGGTAATAG
GTTAAGGAATACCATAAGGGGAAAGGAGACTTGTCCTACAAGTTTCTTAGCACTTAGCAA
ATGGAGCAGGCATTTGCTATGCATTAAAAAAATAAGCATCATCCAAACTCTCAGACTCATC
CAGCCACAAACTTAACTTTTTGTTCCTCCTCCCAGATAAAATTCTCGACTTATTTCC
ATTTGTCATCTTTTTCTCACTAACCGCCACCTCCACTGATGTCTCAGCCCACTTCAGTGT
AGCTTCAGCTTTCATCATTACAGTGAAACAGCTTACATGAAAGTTACCAATGATTTCTAA
AGAATATATATTTTTAAAGTTTATTTATTGATCTTTTGGCAGCATTAAGCAATGTTGTTT
GTGGTTTCATTGCTCATATACTTTCTTCCTACTTTGATTTGAATACTTTTTGCTTTGAAT
ACTTACCTTTCCTTCCCTGACCAGTAAATGCCACTTTGCTAGGTCTCTTCACAGCTCCAT
GCTTTTTTTCAGGTAGTCCCTTGCCCAGGTACTTTTTAAGTGAGGTGAGTATCAAATATA
TATACACATCAGACTAGTCCTCTGGGATACACACAATCACAAATACACTTAAACACTCAA
TGTACCTTTATTATAAATCTTGAAATGAGTTTTTATAAGTCTTGCAACCAAAGTTTAAAA
AAGAATAAATTCTTTTTTTAAATTGCTTTGGCTATTCCAGGTCTTTTGCACTTTCATAAA
AAATTAAAATTAGTACTTTCATTTCCAGAAAAAAGACTGTCGTGGTATTGAACGTGATTA
ATTGCATTAACTCTATAGATCAATTTGGGGAGAATTGCCATATTAACAATACTAAGCCTT
TTAATGCATGTCCACAATGAATATATTTATTTAGGAGTTCTTTATTATCTCTCTGCAATG
TTTCATCATTTTCAGTATATATATATATATATAATGAAATATATATACTTATACATATATTT
TATGAAATATATATACTTATATATATATTTTATGAAATATATATACTTATATATATATTT
TATGAAATATATATACTTATATATATATTTTATGAAATATATATACTTATATATATATTT
TATGAAATGTATGCCTAAAACACATTCTTTTGATATTGAAACTTTTAAAATTTAATTTTC
CATTTATTGCTAGTATGTAGAAGTATAATTGATTTTTTTGTTTATTGATTTAATGACCTG
CTCCTTGCTAAATTCTTTTATAAGTTCTAGTGGGTTTTTGGTAGATTCTTTAGGATGATC
TTTGTAAGCAATAATTTCTTCTCAATAGAACCAATCTGTAGGCATTTTATTTATTTTTCT
TTTCTTCTTGTATTGGCTCAAAGTCCAGTACAATGTTGAGTACGAGTGGTGAGAGAAGAC
TTGATTTTTTGAGTGGTAAGCCAACACTGCATTGCTAGAATAAATCTGATTGAGCAAATG
GTATTATCCTATTTATATATTGCAGGATTTAATTTGATAACATATTTTTAAAGAGATTTT
TATCTCTATTCATGAAGGATATTTAGTTGTTAGCTGTCTTTTGTTGCCATATCTTTGATT
ACAAAGATAAATGTGACCTCATGAAATTTGTTGGAACATATTATATTTTCTGTACATTTC
ATTAAAAGTCTGAACAAGATTGGAATTATTTATTCTTTACTGTTTGATTGAATTCATTAA
TGGACCTATCTGGGCCTGGAGATTTTCTTGTAGCATAGTTTGTAAGTACAGAGTCAGTTT
TGGTCATCTTTGTCTCTCAAGGGCTTTGTCCATTTCATGTAAGTTGGCAAATTCATTGTT
TATCCATAATGTTTTTAATGTTTGTAGCATGTTTGCCTCTTCCTCATAACTTTATCCTGG
TCACAAACATTTTTTAAGACAGAGTAGGTTTTAAGGTCCATCATGTACATGCTATTTCCA
ATTCATAACTGTGGTAATACATTTTTCAGGGTGTATTTTTGCATTAAATATGATTTATAA
AGTTTATTCATAATAGTGAAATAAAAGTGGGGTGCATGTATTTTACTTAATCCTTCTCAG
TGCCTGCTTGATTGAAACCTCTGAGATTTACAATAATGTACTTTTAGGGATGCATTAAGG
ATTACTAGTGCATAGTTCCTGGAGCTCAGTAATGTCAGTTATTCCTCTTAATTTTATACG
GAGTTTCTCTGAATTCTCCATGTCTCTAGACAGCTTATCAATGGAGAAATTTATGTGTCC

FIG. 1AK

TCAAAATGAATGCAGGATTCAGCATCTTCTATCCTTATTTAGATCATTATCTAAAAAGGG
CATCACTACATTTTTTTTTCCCGATTTCAGGGACCATAGCTTTCTCTTTATGAAAACTGT
ATTTTTTTTTTTTTTTTTGAGATGGAGTTTTGCTCTTGTTGCCCAGGCTGGAGTGTAATT
GTGTGATCTCATCGGCTCATTGCAACATCCACCTCCTGGGTTCAAGCGATTCTCCTGCCT
CAGCCTCCTGAGTAGCTGGGATTACAGGCATGTGCCACCACGCCCAGCTAATTTTGTATT
TTTAGTAGAGGCGGGGTTTCTTCATGTTGGTCAGGCTGGTCATGAACTCCCAACCTCAGG
TGACCGAAAACTGTTTCTAATGGCGGCAGAAGTCATCAGATGCAGAATGTAGATTCTCTC
CTTCAGGGGAACAGTCAGTGATAGAATCACTAAAATTTAATTGATCTATCAGAGATCATT
TAGAAGACAGACAGTTCAAGATCATTTAGCAGACACATACAGGCTTTTCATGATAGGAGT
CTCCTGGAACATTCCAGCATCCATTGCTCATTCTTTTCAGTTATTTTTTAAAATTGCTTT
TTAAAATGAGAGTCACAGAAGAGAAAGTTCCTATCTCTCCCCAACCAGTGGGTTAAAAGA
TTGAGTTGAACCACTACTATGTAAAAAAGATTGTCTACATGACAAGACATACAGAGTGAG
AAGAAAAAATAATTTATCCGATATTTTCCATTCAAGGGCAGGTCTTTGTTAACATCATTTG
CCTCTTCAAGAAAGAAAATGGTCAAAGGAAATGTCATATTAATTTATCTGTGTGGACATA
TAAGTAAAATTCTGTTCTCAAATTAAAGATTATCGAACAGACTTTGATCTGGTGGTGTAA
AAATCAACAAAATCTATCGAACATCTATTCTGAGAAACCACAAGGACACATTGGTCAGTA
CTGGTTTGCCGCACAGAGACAGAAAGTAAAAGCTGAATAATCTTAACAGAGCTAAGGTGG
CCTTTTCTTGTGTTTGTTGGCACATTTTCCTCTTTAAAAAATTATGCATGCTGAATTTTA
TTGTTCTGTTCACTAAACCCTATCAATCTTCATGAGCTTACAATTAAGAGAATATTGTAC
TTGGAGGGATTCCCTGCTATTATCAAATAACTTTGAAAAGAAATGGAAAAGTACAAGTTG
TGTAATTACTGTTACAAATTCCAGCTATTTGAAATATTAATGTAAGACCGCAAAAAATCC
TCAATGGGTTTGTGTGCATTTTAAAGGGCTGGACCACAAAACTGATTTCAAACAATTTCA
TAACTACAAATAGTGAACAACAAAAAATACTTGATTTTTTAAAAATATCCTTTCATTCAA
AGGTTTGCTTTGTCCGAACTCGAGAAGCAGAAAACCTGAAAAGCTACAGGTAGTTAAGTT
CTATCTCTCTGGCAGCAGATGGCAGTATTGATGCGTGAAAAATCCATAACAGGTTGCTTG
ACGTTACTTGCTGGGTTTTCCTCTGCTTTAAACTTTGGTATCTGAGCTGAACAAAAATTC
CTAATAAGATAATATGGCTGACATCCCTTTATCATTCTCCTTTCCCAAGCTTTGTTCTTT
TTACAAGGAAATATCTTTTCCACTTGCAGCTTTCTTTAGACATTGACAAAATTTTGATGT
TTTAACTTTTTTTTCCACACAAACTCCTATTTGGTATTCGTCTGAATTAACGCCAAGCAC
ATACTAAGGTCAGCAAATGCTCTGGAGAAACAGGCGCTCAAACCTCCCACACCTCAGGCG
TCTGGAAGCCTTTCCTTACTGTGTTTTCTAATTACTTCCCCAAAGTGGAACTTTCCTAAG
TCAAATTGCAATAAGGGTCTGTCTCTTCTCCTTTCAGATCCCTGGAACATCATCTGTAGT
TCAGAGAAAATGGAAGCCCCTGCAGCCTGTTTCACAGCCTCGAGGGCCAGGACAGCCAAC
GAAGTCCCGGATGAGCGCTGTGGCGGCTGAAATAAAGCAGATCCGAGCCAGAAGGAAAAC
AGCCCGGATGTTGATGGTTGTGCTTTTGGTATTTGCAATTTGCTATCTACCAATTAGCAT
CCTCAATGTGCTAAAGAGGTAAAACTTATCTGTTATTTGAAAATGAAATAGCCTGCCTTT
TCTTGATTCTTAATTAACTTTTTTTTTTTTTTTTAACTAAGCCAGAGAAAAATCTAAACT
TTCTGCTTAGATACCTTGTCAGGCCAGATGACTCAGTTATGTTGTTACCAGCAGGTAAGG
CGAACAGCCTTTAAGAGTGCTCAGACATGTGCTTTTGTCATGCGTATTCTCAGTTGCATG
GCAGACATAAAACAGATGTTTCTCCAATCTCTTCAAGCTAGTTGCTAAACCTTAGATGCA
GACAAAGTTCTAATGCGTAACAACTCATTTACAGCTTGCAGTTCTTTCTTGATGAGAACA
AACGGGTTTTTCAAAACTTCGTTTCCAAAAACATAGGCAATTGTGAGAGAATTATATCTT
AAGGATAAAAAGAGATAAGAACCTTATGTTAGTATTCTAATTATACTTAAAAGTGCATTG
GCGAGCACTTTTTAAAAAAAGCCATCAAGGCAGATATGTATGTCGAATGTCTAAACAGAA
GAATTCATTTCCTGAAGTCACTGAATGAAGCTCAGGGCAGATAGTAATAAAAATCAATGA
GGGAAAGTATGCTATTTGCTACAATGCAGGCACAACTATTAAGTTAAAAATTTTGACCCA
TGACATGAGCAGCAGATGCAGAGGCAGTGGTACACACTAAACTTCATGGCCAGCCAACTT
TATTGGGAATTATCGAAACTGTTCCACATAGACTGGTCCCAAGGCAATACCAATTCTTGT
TTACAACAGGCTTCGAACTTAAGCTAGAATTGCTCCTCTCACTTTGGCCTGATTCAGAAT
CAATATTATATTCCTCACAGCTGGGAACTCTGAAGAACAGCAGCTTTGGCTGGAGTCAGA
AGAAGTGGTATAATCAGCCGCAAAGGGTTCTCATTCTCTTTGGCCCCTGTTTTTGATGGT
TTAACGGCTTTTTCAATGGAGAAGAAATGGAGAACAAACTTCTGTTCAGTGATACTATTT
ACTACAGTCACAGCCTTAAAGATATATGATTTTTTTTGTGGCCTGGGTCCTTGAGACATA
TGCCAGCTCACAAGAAAATAGAAATTATTTTGCCCTCCATATCTTTTTGTGCTTTGCTTC
TTATTAATTATTATTATTATTATTATTATTATTATTGCTAACAGAAATTTAATACAAATT
TATTATGGCCACCTGCTAGATACTCTCTTCTGCTTTTAAGAACTTTATAAACATGTTATG
TTACTGGATGGGTTTTATTCTTTCTTTCTCTTTCTAATCTTTTTTCTCTTTTAGTCAGTC

FIG. 1AL

TAAATTCAATGAAAATTGGATTTACCTCTCTCTGGGTTACGTATTTGGATTTCCTGTATC
TCAGAACTTGCTCTTTCCTTTTATTTCCAACTACTTTTTTTGTCAATAATATATGTCTTA
TTCTTCTGACTCAAATCCCTTCCTTTCCAAGGAAAACAAATAACTTTCAAAGGAGCAAGG
CTGTGTTAAATTTAAGATATTTCAAGTTTTGGGGCATTCTACTCTTTTCCACAACATAAA
AACTTTTGAAAAAAAGAACTTGAAAATTGTTTTTCTTGTTACCACACACATTTTTTTTCA
AATGCTTATATTTATTTATCCTTGCATATGAAATTTGTTTTTTCTTTCTCCCACAAAAAC
CATTCCTAGCTTTTTCTCCTTAAAACTTAACTTTTTGCCAAATTAGTCAAAAGCAATTTC
TTTACAACAGTTCAGGTTTGTCCAAGATTTCAAAGACATTTTGAGGTAAAGGGTCATAAC
ATAGTACAAATTTCTTTTGTCCGTATTATTTCACTCTATATAGTATTTTTGTAAAACTCT
AGTACTCTTTATACCAGAAATGGTATAAGGTACACCTTATACCAGAAATGCATTGTTGTC
ATGCCTTCTTGCTGTAACACACACACACACACACACACACACACACACACACACACACAC
TGATACAGACAATCGAAGATGGAGACTTTAGCAGGAAATATATATGATTTTTGGCCTATA
ATCTTATTGAGTAGCTGTAAGTTATCTGTTATAGGTTAGTGATTAGAATTTATAGATGGA
ATATTTCTAAGTATGGAGAAAATTTTTTAATAGTCTTTAAGGATAGCATAACAAAACATT
TTTTAAAGTTTAAAATAATACATGAAAAATTAACACTCATTAATTTTTAAAAATTACCAA
AATTCTGCCCATCGAGAACTGTTTCTTCTCTGGGTATTAAGGAGTCCCAGAAGGCAAGTT
TCAGATAGTCCAGGAAGATTGGAGTTGAAGGCATATGATACTTTGATCAATACATAAATG
AAAGTAGGAAGAAGTACTTGAAGACTATCATTTAGGAGTGATTTTTAAATGATACACATA
ATAGAATATTAATACAACATTATTCAATTGTATTTAGAAAGAAAATGAAATAAAGAAGAT
ATATATTCAACTTCCATATTCTTATTTACAAGAAATACTGCATCTGCTATTTGTGGGAGG
GAGAGACCCTTCTCTGACCTGATTTGGCTTTTGATTTATTGATTGTGCTGTGGAGGTTCT
GTTCAGGCACTGAAGTTATTCTAAACCAATTATGGGGTCAGAAACCAATCTGTGGTCAAT
TCCTGCAACTGAAGAGGACAGGAGTCAGACCATCCTCTACCAATAGCCTTGTTCACCTTT
GAATTTAATTATTTAAAAGACACTTTTCTGTTGTTTCTTTTCCTGCAGAGTATTTGGGAT
GTTTGCCCATACTGAAGACAGAGAGACTGTGTATGCCTGGTTTACCTTTTCACACTGGCT
TGTATATGCCAATAGTGCTGCGAATCCAATTATTTATAATTTTCTCAGTGGTGAGTTTTC
AACTGTTCTTCCATAAGCCACAATTGTAACCAAGGATGAGGAATCAATGAACACTCTTCA
ACTATATGAGGAGTTTAGTTGCTATGTGAGTTGTATTTTTTCCCTGACCTGATTTATCTT
GAGTTTCTTCTCTTTTGAGGCAAAGTATTTGTTACTGAACTCATCAGAGAAAATGAACTG
ATTTTTCCATGTCAAACGTATAAGAAATGTTATAATAGAAGAAAAGTAAACATTCTGAGA
AATCAATAACACAAAATCTTACATGACATACTTTAAACTCATGATTTACAAAAATATAAA
ATACTTTGTTCTGTTTTGCCTTGCTATATTATTCCTTTGCCAAAATGTGTAGCCTAATTG
AGACAGAATTGGGATCTATTCACTTTTAGATATTTTACTATATTTACGTTTCTCTTGTGA
GTATCATCTCTTGGATTTATCTCAATATTTCCCACTGACTACCAAAAATAGTATTACTCC
AAAATAACACATAAGTTAAATGATACACACATACATATACGTGTAACTTATACAATTTGT
ATCTGTTTATGGAATCAATATAATTATAAAAGTCATTTAAATCACTATTGTTTATTCACA
TTTTGCCCGACTGACTTTTAGAATTATTTTTAATTAGCTACCTTTTTACATTGCCTTAAT
CTCCAACTCATTGGCGATTTCTTTGTTATTTCTATCTTCAAATATATGGTGATTTTATGT
GGAAGAATAGAAATTCATTTTGTGGCATATTTAATAAAGCTTCTGCATCTTCCAACTTGA
TCTTTGGCCTTCTGGTTTGCATAGGTTTAAAAAAAAAGGCAACAAATTAGATTGATGAGAA
ATAATTTTGTTCTATTTAAAAAAAAATCTAGCACAATGACTAAAGCTCTGAACCTCGCAC
TAAGCAGGTAAAGGCTATGAGGAAGTTGTAATGAGAAGTGTTTGAAGCAGAAGTCACAGA
ACCAGGTCAAAGTCCTAGTATGGAGGATAAAAGTGAGTTAGAGGAGGCAACTGATAATCA
CTGATAACTCATTATGTGACTGCTATTGTGCTGGGCCCGTGAACATTCATCTTCTCATTT
AATCACTGATAACTCAGTGCGTAGCTGAGGCTAAGAGAGAAGAAATGATTCGCAATACTG
CCATTCACACTAATAAAAGTGATTCATTCATTCTCATAGTTCTCCAATATCTCCTCCATA
ATTTAAAGACAAGGAATAGCTTCTACAGTATTTTTCCCCCTTCAGTTTTTGTTCTTTCTT
ATATAGATTATGAAACTGAAAATTTTCTGGATATTTGAGTGTATGTTTCTAGGTATTTTG
TGGATTTAATTGTTTCAGTATCAGTTATTTAGAGTAAAATGCAGGAGTAATTTTTGTATA
ATTTTGGCTTTGTATGACATAAGTTTCATTGTGTTTAATTATTAAATATCTCTGAGAGTT
CTTCTACTGATGATCACTTCCATTATAGTTATGTAGATAAAATATACCAATATGCGTAAA
TATATGAGGTTTGACTATAAAGGAATGAAGCAAATTCCAAGCCCCATATGTGAAAGGCAG
CCTCGTTATTTTATGAAAATATTCATTGTTTCAAGAGTCTACCAAGCTTCCAATAAACTC
AATTTCCTTATTCTATTTTACCCATCTTTGCAAAATATTACACCTCATTGTTAGTTTGGC
TCAAGGGAGCAACTCAGTTGTACCCTATTCATAATTTGTTGAAGCATTTATGTATAATTC
CTTTTCCTTTCATTCTCTCTGTTTGCCAGGAAAATTTCGAGAGGAATTTAAAGCTGCGTT
TTCTTGCTGTTGCCTTGGAGTTCACCATCGCCAGGAGGATCGGCTCACCAGGGGACGAAC

FIG. 1AM

TAGCACAGAGAGCCGGAAGTCCTTGACCACTCAAATCAGCAACTTTGATAACATATCAAA
ACTTTCTGAGCAAGTTGTGCTCACTAGCATAAGCACACTCCCAGCAGCCAATGGAGCAGG
ACCACTTCAAAACTGGTAGAATATTTATTCATATGACAAGGATACCTGAGTAAAACTATC
CTTTTTAAAATCACTGGGAACAGAAATTTTATTATCCTATGATGTGAAGCTAAAATTACT
TGTGGATCTTTTTTTTTTTTAATCTATTGCTCTTTGGAAATAAAAAAAAAGTCAGTTTAA
AATGATTTCTCAACTTTTGATTTAAATATGTTAGAAGTTTAACCTTCAATTGAGCTTATT
TCAGGCTATTTCACTTTTAGTTTCATGTATTAAAATGTGTGTCAATTAAATGTTTAAACA
TTTCTAATTCTTTTTATAATCCCTTGTTATTTTAATCTCTCACATTCAGATTGGTTCCTA
AAAATTACCAGAATCTATCCAATGATTTTTTTGCTACTAAAAGAAGTAGCAATTACTAA
TTCTGAATTAACAATAGACATGTTAGTTGACTTAATAGTTTTTTTAAAAAAATCACAAG
ACTGTTGTTATAATGTGATATCTGAGAAAATATTTATATATAAAATAGCATATTGTGTTA
GGTAATTCAAAAACTATCTTACAAAACTATATACTCACTCATTTACACAATTTTCTCAGG
TTTGCAAATTGACCATTGCTAACATTTCTTGTCTCAACATATGGCCAGTAAGACTCTATC
ACAGTAAAAGTTTTAACGTAATTTCCATCTCTAACACTTTAACATTTAAGAATAAGCTAA
ATCACATCATTATATTCTTTTAACAACAACAACAAAAAGTGATATAGTCAGCCTTGCTGG
ATTAAATTAAAAATGCACCACTGTGCTAGGTGCTAGGGAATGAGATGGCGTCGATGCAAA
CATGCCTTCAAAAGAGCTTCAGTCTAGTGAGGGAGACATGTTGACAGAGTGCAAGGCAGC
AAACAATCTGGGGGACAATTCTTGGTCATGGCAGAGCAGTGAAGCTACCAAGGACAGTGG
TCTTCACCAGAAAGTTTGTAGCGCAGTTGCACTTTTTTTTTTTCTTCAATTTAATTACAA
TGACAGTTGATGTCAGTGGATTCCATCTGGCCTGGGGCTGAGTACCAGGTGGTTAAAAA
ATAGAGGGGCTTGCTCTTAACTCACACATACATGAATAGACTATCGTATATTTTGTAGAA
AATGTAAGATCTGGGAGTCAAAGCACTGAGTATTCAAACTTATTCCCCTGAAAAATTCTT
CTGATTCAAATATTTACTTGAAAATTAAACTAAAAGTAAAAGAAGTGTTTATGAAAGATG
ATTTTCATCTCCTATTATGGTAACAGGTGTTCTGATTGTATTGAAACAAAAGATATGGGG
CACAGTGTTTAAGAAAAACTTTCATAGAAAATTAATTTTTGTTATTTTTTCATTTTTCCA
TTACACTCAGAGAAAAGTAAAAGAGCCTAATTATCCACAACTTGTTTTCAAATCTTGGAA
TTTGGGATTCTGTTACCTTGTGCCTTTTATGACTCAAAGCAAAAACTATCTTCTTATACA
AGGTTTATTGAGATCATATTGTAAAATATCAGCACTATATCAGTGAAAGCAAGGTATTTT
AACTCTCTCCTTTCTCCCCTTTGTATATTTTAGAACTTCTTTTCTTATTTATGCTGCATG
AAAGGAGAGTTGTAATTTTCGGATCGTGATGACAGCACTTTAAAAAGTTTGAGGATAACT
TCAAATAACGTTGATAATATGCCTTAATAGCCAGTAATAGCTCAGAGGAAGAGTAAATTC
CTTAGACCCAAATATACCTACTTATAATTTAAAAGAGAGAAAAGGGAGAGAGATTGAGAG
GGAAAGGGGGAGAGAAGGAAAAAACTGACTCAGAGAGCAGCAGTTATGTGACGTATGGGA
AGTCAGAATTCCTTTGCTCTAAATCAGTGATTCTCAAAATGTGGTTTCTATACCATCGGC
ATCAGCATCATCTGGGAACTTAGTGGACATGCTAACTTCCACCCTATCCCTCACCTACTT
AACCAGAAACTTTAGGGGTGATAGCCCAAAAGCTGTGTGTTAAGCACTACAGGTGTTTCT
GAAGCACTTTAAGATTTGAGATCCACTGCTTTAAGTGATACCATCTGACATCAGTTTATC
TGCCTGTGTGAAATAAAGTCTTTTACTGCACAGGTGTCTACAACAGGGGCCACCATCATC
GCTACCGTCAACGTGGTTGGATGTCTGAAAGAAGAAGCTGAGTATCAATGTTGACTCTCA
CTCATGTCATCTTATTAAAAAAAAAACAGTTTACAAAACAATTGCTACTGATAAATGCAG
TGTGAAAGACTGGTTTTAAGGCACTTGTGTGCTTTATGTCCACCCAGATAACTTGAGTTT
TTAACTAAAAGTTTCAAATCCCTATCTTCTTTATACTTACTAAAGTTCGTTTTGCAGAAG
CAGATAGTTTCTAAGAATGATCATTTCATGGAAGGAGATATAAAATAAAATAAAACCAGT
ACTTAAACTCTGGGAATGTAATAGGCCATGTACATAGCACTCAACATGTGAATCCAGGAA
TCCTTCTAAGAGGTCTAGATTTAGTATGGTTACCTTAATAGGACAAATGGTAAAGAAATA
GGTGTTCCCAAACTCTGCCAATCTTATGAAACAAAGAGTCAACTCTTTACCTCATTATTT
GCTAATGACACAAATGCAAAGACATCTTTTGAAAAGAATGTGTTGGGACTGTTTTATGCT
GTACCTTGAATGTGTATCTCTCCTTTTTTTGCTATATTTCAAAGATTTAATGTAAGTTGT
CAATGTCATTGAGTTCTTGTTATCAATAGGGATGATATAATTTTATCTAACATGGAATCC
ATTTTAACTTTGTTATTTCTGAATTTCTATGAAACCACAAAAACCTTCATACTTGAATTT
ATTTATTCTTGGCTAAAGATTACTCCCAGTTTGTGAGGAAATTTATTTCTGAGTTTCCAA
AGCTTGGAGAATTTATTGGATATTAAATACCTGTTATAAATTATTGATGAGTTAATTGCA
AGTAGCAGACACAATGATATTGAATTTCACTCCCAATACACATTGTTTTAATGAAGATTA
AGGTAAATATGTTTATAAAATTTAGTCTGGCTATGCTTAAACCTGAAATAGCAGAATGGC
AAAAAACCCCAAAGCTGTTTATGGACCCAAATTGTGAGGAGGGCTATTATTTTAATACTT
GTGTAATAATAGAATGCACTTGATGTAAATTGTAATAGCCATCAACTGCATTTCAAAAAC
CTTTCGCTAGCCACTACAATTTAGAAAGCTTTTCAGTGTCAGTTAGTTTTACAACAAATG

FIG. 1AN

CCTTTTCTACTTTCTACAAGTCACAAGTCAAAAAAAAGTAAATTCCACCAAGTTTTATTC
AATTAGTTTTCAAATTGCATGAAGCAAAAAATAGATTTTAGAGACAATATATAAATAGA
AAAAATATTGTAAAAGTCTACTCTATTACCTTATGTACCACAAAAAAATAAAGTACAAAG
GCATGAAAAACACTATTATTTCCCAAAGTTCAAAGGGAATTGTTTTCTACGCAACTACTG
CTACTAACAAGGGGACAACAACCCCCTCCACTTGCCACGTATTTTTATTCTCTTTTCTTT
ATATCTTTGGAGTTAAATGTCTTTTATGTTTTTCATGAAATGTATTCTATAATTGTTGTA
TTTCATGTGTGTAACATTATGTCAGTTGTTTTAACAATTATCTTATATCTTGAAATTCTT
TATGCCTGATTGTACTGTGTCTTCATGAAGAAATTTCTTATCAAATCCAATGTGATTACA
CACTTACTGCTGTAAAGGATGCGCATTATGTAGTTTTTAAGTAAAAACTATAGTGAGAAT
TCTATAATCACATTCACACTCCCCTCTCTATTGTATGAAAAATCTTGTTGTTGTTGATTA
GATAAGGTGGATATTCACTCATAGTTAATGTCAAATCTCTGCAGTTAAGGATTGAATTAA
GCCCTCTGGTGCAGTACCTAATGATCAAAACATTTTTTCCAATAAGTTTATATAACCAAG
GATAATAATGATATAAAAGGTTTTTAATGTTGTTTTTAAGAGCAGGTACTATAACAAAGA
AGGTTAACACTGGTACAGAAATATTTCATAAAAGTTATGAAAACCAGATAAATACAGTAT
TAAATTTTGGAGCTTTTATCTGAGTTGAGAGATTTAGTCTACATTGACTGAGATGAAATG
ATGAACTCATAATTTTCAATTTATTATCAGAATAATAAGTGACATTTACATAATTAATTT
TTTTCTGGGCCATTTTGTATAAGTCATTTAGGACTATTTTAAGTTCACTGGTAAATTTTA
AAATGTATATTTTCAGCTTTTCAATTTTTTTCAAAATAGTTCTGAGAAATTACAGAATCA
GATACTAAGGATATTAATTTAAAAATCAATTTTTATTCAGCACTATTTATTCTAACATAT
ATAAAAAAATGAAGCCAAAGTAACCCGTCAAGGTAAATACTTGACTCCTAGGAAAATGTGA
TTTTAGTAGGCATCTCAAGAGGAAGTGAAACTTCTCGTGGTGAAATTACAAGAAAAACAA
GTTATTCAGTGGTGAGAATGTGTTGCTCTAAGCAATCCATTAGCACAGACTAGCTACTTG
GCCACTCCTCTTCCTTCTGGAGCCAGCCCTGAAGAGTGGTCACAGCATCTTCATTTTTAT
CCAGGCCAATGGCCATGCATGAGAAGTTGGGTAGCAAAATTCTTGAAGCACCTCTTTGTT
CTTGCTCTTCTTTCACTGTTTTCTCACTCTCCACCTGTAATGCTCACTGCCAGTTTTACC
ACCAAGCTAAGTATCAGCAGACCTCCCTCCACAGCGTGCCTTGCCCTGTAGAACTCCTGG
TCCTTCCTTCAGCCCAACCCCATCCAATTGCCTAGGTTCTTGTTGTCTCCTGAGATGAAC
AAGAGGCAAGTAGCTAATTTGAGAACAAATGAAGCAGAGCTGAAGGAAAAAGTAAAACAT
TTATTTTTTCATATCCCAAATTTTATAATTTTACATTTTTTTAAAACCCATTCATTTTCT
TCCCAGAACATTTATGCTTATCAGTGGTCTTCTGAATCTGTGACAACTCCCTTTTCAAGC
CCCAGCTAAGCTTCTTGCCTCAAGCCAGAAGGAATCCCAGTTTTGAGTCTTGTGTTAAGG
CCATGGCAGGTCAGTAGGGAGATTATCTGAGGAGGTACCGCTTGTGACACCTTCAGAAAC
AAAACAGCTATTGCCTTACGTTTCATAGGCCCAGGCCCTGAGCAATAGCAAAAAGATAAT
ACTTATTTTTTTAAACTGTTGTTTATTAGGTGATCGATTTCTAATTAATTTCAAAATATT
TAAGGTAATATTTTAATTACCGAGGAAGAATGGTACAAACAAAATGTTGTGGAACTGGAA
AATCCTCAGTGCTTGACAACATGAAACTTATTTAACTTATTATAGATGAGATAATGAGAA
CATCTTCAGAAAAGAAGCTATGTTCCTTAAAACAGGGGTACAGATTTAAAAGCTCTGTTT
ATATGGTTTTGGTAGACTAAGTGAAGAACTTGCCTATAAAGCTGAGTCTCGATCATATAG
CATATCCATTATAAAGTGAGAAAATTGCAATTTTAGAGTATTGTCAATACATCCAAAAAT
TTTTACATGATTTCTAAATGCAGATGTGTGTGTGTGTATGTCTACGTATGTCTCTCCATA
TGCAACAAGCAGTTAATTAGTCCAAATATATCCCACAGTGTAGATTAGTTTCATATCTCA
GCTCTTCAATGTCTCTTCTTCATTTAATTCACTCCTTGGTGTCTAGTTTTCCTCACTCTT
TTACAAATATCCAGGTTCTATATTTCTGCTTTTCTAGAGAGCTTTTTCCCTCAAGAATAT
ATTTTTCTTTTTCTTTCTTTCTCTTTATTTTTGTTTGTTTTTAACTAACATTCATATGGT
TATAACAATTTGAGACAAGTGAAAAGGAAAGATCTGTAAACTGCCTATCTCCTTTGAAAT
TCATTGCCAATAATCCTTAAGAATATAAAGTTCCTTGATGCCAAAGACCTTCTCATTAGT
GTTGCTGCCTGTTGTTTCATTGGTTCCCTAGAACAATGCCTGGCACATAAAAGTTATTTG
ATAAATATCTCTGCTATTAATGAATTAATAATAACTGCATGACAATTCTTTCCTCAATTC
ATCATTTTGCTTCATTTTCTCACAGTTGCTTCAATGTGTCTGTGGAACTATCTTTCCATG
TGAACAAAACACTCTACATTCTCAGTGTCTACAAAGCACATATTTCCTTTTATTAAAATT
AAACTTTGAGAGCACCAAATCCTAATGTCTAACCATCATCAAACTGGCAGATAGCACCAG
TATTCTTTTTGCTCACCAATTTTATGCCCAGGCATCTACTGTTTCTTTCATGAATAAAAC
CTGACACCTGTAAGAGGATTTATCATGGTAAACTTCTCTTTGTTACTGACATTTTCAGCC
TCTTGGGCTCTCCCCTCCTTACTTATACACATTGGCACCCAGCTTGAAGTCATACTCTCT
AGACCCTGGGTCAATGTGGGTAATGCATCCAGGAATCCAGCTTAACTCTTCCTTGGTCTC
TTTGATGTGACTGACCTTTATTTCTACATTTCTTCATCAAACCAGTCTCACAGTTTTGCA
CAGTGCAAATCACATGCTGCACCATGTGCTTATTATCTCCTATAACAACAGATGCTCCAC

FIG. 1AO

TGAAATGCAAAACTCTGTGTTAAGCCAACAACTGCTTCTCCATCCTTTCCTCCTATACGT
TTCTTCTCACTACAACTTCCCTTCTCAACCCCAAAGGGACTACTGGATTCTTTACTCTTT
TATTTTACCCCAGTCTATCAGTCCCATCCTGGACTTCCTTCCTTCTCTGCTTAGAGGAAA
GCAAGATGATCAGGTAGAATTGCACTTATGACTAGATATTATTTACTTCAAACAAATTCT
TACTATTTTGTCCTGATGAAATTCATGACAGTTTTCATACAACAGAAAGCCTGCCCTCTT
AGAAGAGAAGAGAACTGAAAAGAAATGGTTGAAGTAAGGTAGAAAGCCCTCATGGAGTTA
GGTGGCTAGGCCAGCAGAGCTAGGCACTGTTCTCCTGTTCAGAATTGCACTCCTGATACT
CCAGATGGGAAGCCTGCCATGGCACTAACCACAGCACTTTTTATACCCTATCTCTGCTAT
TATGAGCCCACATTAGTTTTTCTTCTGCTTCAGAAATTGTTGCAAAAAATAATTTTATTA
TTTACAAATTATTTTTAAACCATATAAATCTGCTTAGTTTGATTTCTCAAACCCTCTAAA
ACTTACACTTCTTGTTGTCCAATCTTTGCTTTTAATTGGGTATAATTTGAGGCAGAAATA
AATTAATCTCATTTTTAAAAATGTACTAGCTATTAATAATTTTTAAATTTATCTTCTAAA
ATTGGAAAGTATCCACTTTAAATGCATCTGTAGCAAGGACTTTTTACATACATTCTGTAG
CTTTATTACTTCCATTGAGAACTGTTAAAATAACAGAACTTACCTCACTGTACGCTGGCT
TTTGAAAAGGCAGCAGAACTGTTTATCTGATTATCGAAGTAATCATATTACATTTCTTTT
TCTTTTCTAAGAGAAACCTTCTTCATGTGCTCAGTCAAACATTTTGGTGTTTAAGAATTG
ACTTATTAGGTCAGGCGCGGTTGCTCACGCCTGTAATCCCAACACTTTGGAAGGCCGAGG
CAGGTGGATCACTTAAGGTCAGGAGTTCGAGACCAGCCTGGTCAACATGGTGAAACCCCA
CCCCTACTAAAAATGCAAAAAAAAAAAAATAGCAAGGTGTGGTGGTGCACATCTGTAATC
CCAGCTACTTGAGAGGCTGAGGTGGGAGAATCATTTGAACTCGGGAGGCGGAGGTTGCAG
TGAGCAGAGATCACACGACTGCACTCCAGCCTGGGCGACAAACAAGAATCTGTCTCAAAA
AAAACACAAAAAACAAAAAACAAACAAAAAAGAGTTGACTTAGTTAATGAAAATATTTTT
ATTAGGAAATTATACTTCTCTTTACAAAGTATGTATTATTTGTTGCATCTATATAGTCTA
TCAATTCTAAAAGCACACTTTATGCGAAAATGTAGTCTAGGCCTTCAGAATGTATTATTA
CAAGAAAGTATCTATCAACCATGTTTCATTTGTTTGCATGTTTTGTTTTGTTTCCAATAG
ACTATGAATATTCAGCTTCAAATGCTACCTCATGATTGTTACATTCCTGTTGTTGAAAGA
ACCCATCTCTTTCTTACCTTCTTGTCCCTAAATTGTGTTCTTCTTATAACTTACTTTGCA
CATAACCATAATGGAGTGAGATCATAGAATTAAGAGGATTGAGAAAGAAAAATACTTCCC
TCATTCCATTGGCAGTAATCTGTGATTCAAAAGTTAACAACATACCATGTATTCTTGTAG
GAGATTATTTCATGCTTATCACTGATCAACTTACATGCAGGTTAAAACCAGCCCTGAAAA
AATGCTCATCATCACTGGCCATCAGAGAAATACAAATCAAAACCACAATGAGATACCATC
TCACACCAGAAGAATGGCGATCATTAAAAAGTCAGGAAATAACACTTGCTGGAGAGGATG
TGGAGAAATATGAACACTTTTACACTGTTGCTGGGAGCGTAAACTAGTTCAACCATTGTG
GAAGACAGTGTGGCAATTACTCAAGGATCTAGACTAGAAATACCATTTGACCCAGCCATC
CCATTACAGGGTATATACCCAAAAGATTATAAATCATGCTACTATAAAGGCACATGCACA
CATATGTTTATTGCGGCACTATTCACAATAGCAAAGACTTGGAACCAACCCAAATATCCA
TCAATGATAGACTGGATTAAGAAAATGTGGCACATATACACCATGGAATACTATGCAGCC
ATCTGAGCAAACTATCGCAAGGACAGAAAACCAAACTCCGCATGTTCTCACTCATAGGTG
GGAATTGAACAATGAGAACACTTGGACACAGGGTGGGAAACATCACACACCGGGGCCTAT
CATGGGGTGGGGGTAGGAGGGAGGGATAGCATTAGGAGAAATACCTAATGTAAATGATGA
GTTAATGGGTGCAGTACTCCAACATGGCACATGTATACATATGTAACAAACCTGCACGTT
GTGCACGTGTACCCTAGAACTTAAAATATAATTTAAAAAAAAAAGCCCTAAATGCAACTT
GTTCAGATAACTGGAGCCATCTTCCTAGCTCTTTATTTCTCAGACAGTGTGGGTAAGTCC
TGCTCCGTACGAATGCTTATGTCAGTTTTGAAGTTCAGTACTTTCTTAAGAGCCAGAGTC
AGTCAAGATGTTCCCTTAACAAGATTTTTCAATGGGGTTACACATTAATGAGTTCTTTTT
CCTCCTTTAAGTATTTGAAAATTTTGGTTTAATAAAAGGTTTAACTATGATGAATTTAGG
ATCCTTTTTCCTGTTACAGAGCACAGAATAATAGTTAATATTTTACATACATATTGCAAG
TTCATGTTGCCACTAGGAGTGTCCAGAATAGACAATTGAAACAGCCTTCTAGCTACTACT
ATCAAAAAAGAGCTTTAAATAACATATTTTAATTAAATAACATTATTTTCTATAGCTATA
CCTCAATAAAACCATCAACCAATGTTTGTACAATTTGATGCCCCCACTCTAAGATTTTTA
GCTAGTGTAAATCAGAGTCTCCTATTTAATGAGACACTTTATCCAATCAGGTTGTGTTTA
TTATTCAACCAGATGATCTTGGAACTTATAACAAACTAGTAATACTTAAAGCTGGGCTTT
ATGTGCGTGATTTACTGGGATGTTTGCTTATACCTTGTTTCCAAGCTAAAAATATTGTGA
CCAGGTGTGTTAGTCTGTTTTGAGTTGCTATATAGGACTACCTAATGCTGGGTCATTTAT
AAAGAAAAGAGGTTTATTTGGATTATGGTTCTGCAGGCTGTACAAAGAGCATGACATCAG
CATCTGCTTCTGGTAATGCCCTCAGGAAGCTTTTACTCATGCCAGAAGGCAAGGGGAGCC
AGCGTGCCACATGGCAAGAGAGGGAGGAAGCACAAGAGAGAGGAGACGTACCAGGCACTT

FIG. 1AP

TTTAACAACCAGCTCTCACATGAACTAACAGAGTGATAACTCACTCAATACCCGGGGGAT
GGCACCAAGCCATTCATGAAGGATTTGCCCTCATGACCAAATACCTCCCACTAGGCCCAA
CCTCCAACACTGGGGGTCTCATTTCAACATGAGATTTGGAAGGGACGACTATCCAAACTA
TATCATCAGGATTTTCTGGCATGGACTACCAAGCCATTTCTGCTTCAAACTCCCCTGAAA
TTCTTGTTAAAAATGCAGATTCTTTGATACCACCCCCAATACACTATTTAGTCTGAGATG
AAACTCAAGGATTCTGATTTAATTGATCTAGACTAGCATTTGACCATTGATTTATCATCT
GGGATTCTAGGAAGTCAACCACTTATATGTTTTAGAGCAGACTTCATTATAATTGAGGAG
AATGTTTGTAGTCTGTGGGCTCCTCTGTCCACTTCTGATTGGGGCCCCTTTGCCTGATTC
TGACTGGATCAGGCAGAGTTTTATTCAAGCCACTGTCCTTTTTGGCTTCTTAATGTTCAA
AATATATTAACACAATCTCAGTTTTCTAAGAGCTAAATTATACGACTTGGTTCTTGTCTG
GTAACATAACTGCATTACTGGATCTTGTCAAGATTCAGAGACATTCTCCCAGTTTCAAAT
TTGTAACTAACACTGTTTGATCACAAAAAAGTTCTAAGCCAAAGCAAAACTCTTTCTACC
ACCACCAGATGGCGTTACTTTGGACTTACCTATAAATGGATTTCCAAATGGTTTTTCAGA
AACCAACTGGAGGTACTTAGAAAAACTTATGGAACTCACAACTATTCTTTGCATGTCAAA
AGCTATAACAGTAAATAATATTTGTGGAGAATATTCTGTAAGATTAGGCTGCCTTTCTTT
TCCTCCAGCTTATTTAAACTATATCCTTATATTAACCCTTGTTGGAGATGTGTCCTCTTA
TTGCACTGTATGTGAGTGTGTGTGTGTGTATCCCATCACGTTGGTATGATGATAGCACCC
TTCATTGAGAAGCTTTGCAAAAAGAATATAAGAACATGTTATTATGTTTACTTAAAAGTA
TAAGGCCGGGTGTGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGGGA
GGATGACGAGGTCAGGAGTTAGAGACCAGCCTGACCAACACGGTAAAACCCTGTCTCTAA
TAAAAAATACAAAAATTAGCCAGGTATGATGGCACGCATCTGTAATCCTAGCTACTCAGG
AGGCTGAGGCGGGAGAGTCCCTTGAACCCAGGAGACGGAGTTTGCAGTGAGCCTAGGTGG
CGCCACTGCACTCAGGCCTGGGTGACAGAGTGAGACTCCATCTCAAAAAAAAAGAAAAAAG
AAAAAAAAGTATTGAGGACATTGCTCATGACATTCCAAGGTTATATAAAAGAATATATAA
AAAGAAATTTCTGCCTGGACTTAGTGCCAGGAATACTTGTACTTTTCTTGCTTTCTTCTT
AAGAACATTGCACAATAGAGTATTTTTAAAAATTGTGCTTGCTGTTCAAATTGCCTGCTG
GAAGGATTAGAGGCAGATCTGTAGCATGCCGAGTCCCATCTTTGCATACAGGCTATCATG
ACAAACATTGTATGTGCTAATTCTATCTGGCTTCTCTTTATATTCCTATCTGTCTCTATT
TCCTGTCATTTTAATGTTTTAAAATTGTACTTTTTACTTAAATGGTTTTTGGAAGAAATA
AATATAAGTAAAGTCTGTTAGAGGCCCGGCGCGGTGGCTCACGCCTGTAATCCCAGCACT
TTGGGAGGCCAAGGCGGGTGGATCACAAGGTCAGGAGATTGAGACCACCCTGGCTAACAC
GGTGAAAACCCATCTCTACTAAAAATACAAAAAAAAAAAAAAATTAGCCAGGCGAGGTGG
CGGGTGCCTGTAGTCCCAGCTACTCGAGAGGCTGAGGTGGGAGAATGGCATGAACCCAGG
AGGTGGAGCTCGCAGTGAGCCGAGATCTCACCACTGCACTCCAGCCTGGGCGACAGAGCG
AGACTCCGTCTCAAAAATAAAAAAATAAAATAAAAAAATAAAGTCCGTTACAAAGCACAA
AAAAGAACGGCAAAGCCAACAAACATATGAAAAAAAGCTCATCATCACTGGTCATTAGAG
AAATGCAAATCAAAACCACAATGAGCCATCATCTCACGCCAGTTGGAATGGTGATCATTA
AAAAGTCAGAAAACAACAGATGCTGGAGAGGATGTGGAGAAATAGGAACGCTTTTTACAC
TGTTGGTGGAGGTGTCAATTAGTTCAACCATTGTGGAAAGCAGTGTGGCGATTCCTCAAG
GATCTAGAACCAGAAATACCATTTGACCCAGCAGTCCCATTACTGGGTACATACCCAAAG
GATTATAAATCATTCTACTATAAAGACACATGCACATGTATGTTTTTTGCAGCAGTACTC
ACAATAGCAAAGACTTGGAACCAATCCAAATGCCCATCAGTGATAGACTGGATAAAGAAA
ATGTGGCACATATAATATACAGCATAGAACACTATGCAGCCATAAACAAAGGATGAATTC
ATGTCCTTGGCAGGGACATGGATGAAGCTGGAAACCATCATTCTCAGTAAACTAACACAG
GAACAGAAAACCAAACACCACATGTTCTCACTCATAAGTGGCAGTTGAACAATGAGAACA
CATGGACACAGGGAGGGGAACATTACACATCGGGGCCTATTGGGGAATGGGGGCTAGGGG
AGGGATAGCATTAGGAGAAATACTTAATGTAGATGACGGGTTGATGGGTGCAGCAAACCA
CCATGGCATGTGTATACCTATGTAACAAACCTGCATGTTCTGCTCATGTATCCCAGAACT
TAAAGTATAATAATAAAAAAAAAGAAAGCACAAAAATAAAAGTACTTGGAAAAGTTTAAA
GGGTTAAATATTATGCAAAACTGAAAACTAGCTTCAGATACATTTAAGTTTATATCATGT
TAACAAGTTATTTCTTTCTAAAAAATTCTAACCTGTAACACAGAGAGTGGACTTGAACTT
GAAAATATGGTTAAGGTACAAATGCAGATTTGGGGTCCCAGTCTCCCAGACTGTGGCTTC
TATGGAAGAGATTGTACTGGCTCCAAATTCCACAGATGATTGAACAACTTGTTTCTGCCT
GTGTCAGAGCTGAAGAGTGAATATCTCCACTATATATATCTCAAAATCTCCCAAATGAAA
TTTGGTAACCCTCTATGCCATAACACATCACATTAATAATTTGTATTCAAAAGTCTCTCA
GAAAAGATTTTTGAAATGCCAGATACTTTAATTTTTTTATGTTTATATATTTAGGGTGTA
TGAGTACAGATTTCTTACATGCCTATATTGCATAGTGGTGGAGTCTGGGCTTTTACTGTA

FIG. 1AQ

GTCATCATCTGAACAGTGAACTTGTACCAAATAAGTAATTTTTCAACTCTCATCCACCCA
CCCTCCCATCTTTTGTAGTACCCAAGGTCTATTATCCCACTCTGTATGCCTGTGTACCTA
TTGTTTAGCTTCCACTTATAAGTGAACACATGCAGCATTTGACTTTCTGTTTCTGAGTTA
TTTTACTTAGGATAATGGCCTCCAGTTCCATCTACATGGCTGCAAAAGTTATGATTTTAT
TCTTTTTTATGGCTCCATTATATGTATGTGTGTGTATCTCAATTTTCTTTATCAAACCCT
CTGTTGATGGACACTTAGATTAGTCCACATTTTTGCTATTGTGATAAACATGTAAGTGCA
GGTATCTTTGTAATATAATGATTTCTTTCCCTTTGGATATATACCAGGTAGTGGGATTTC
TGGATCTAATGGTAGTTCTATTTTTAGTTCTTTGAGAAATCTCCATACTGTCTTCCATAA
AGGTTGTACTAGTTTACATTTCCACCAAAAGTGTATAAGCATTCCCTTTTCTCTGCATCC
TCACAAACATCCTTTGCTTATTGACTTTTTAATAACAGCCATTCTGACTAGTGTGAAATA
ATATTTTATTGTGATTTTAATTTTCTCTGATGATTAGTGATGTTGAGCATTGTCTCAACA
TCACTATGCTAGTGGCATGCATGTTTTCTTTTGAAAAAAAGTTTGTGTTCTTTGCCCACA
TTTTAATGGGGTTATTTGTTTTTTTTTTTCTTTGAGTTGTTTGAGTTCCTTGTAGATTCT
GAAAATTATTCCTTTGTCAGCTGCATAGTTTACAATTTTTTTCCCATTCTGTAGTTTGTC
TGTTCACTCTGTTGATTGTTTATTTTTCTGTCCAGAAACTTTAGTTTAAGTCCCATTTGT
CTATTTTTGTTTTTGTTGCATTTGCCTTTGAGGACTAGGTCATAATTTTTTGCCTGGGCA
AATGTCCTGAAGATTTTTTTCCAGGCTTTCTTATAGTATTTTTATAGTTTCGGGTCTTAT
GTTTAGGTCTTTAATCTATCTTGAGTTAATTTTTGTAGCTGGTCAGAGGTAGGTGTCCAG
TTTCAATCTTCTACATATGGCTATCCAGTTTTCCCAGCACCATTTATTGAATAGGGAGTC
ATTTACCCAGTAAATATTTTAGTTGACTTTGTTAAAAATCAGTTGGTTATAGGTGTGTGG
TTTTATTTCTAGGTTCTCTATGCTGTTCTATTCATCAATGTGTACATTTTTATACTAGTA
CCATGTTGTTTTGGTTACTATAGCTTTGTAGCATAATTTGAAGTCATAATATGATGCCAA
CAACTCTGTTCTTTTTGTTTGAAATTGCTTTGGCTTTTTTTCCTTGTGAGAGTTTGCTGA
GAATGATGGTTTCCAGCTTTGTCCATGTCGCTACAAAGGACATAATCTCACCCTTTTTTA
TGGCTGCGTAGTATTCCATGGTGTATATGTGCCACATTTTCTTAATCCAGTCTATCATTG
ATGGGGGGAGGGGGAAGGGATAGCATTAGGAGATATACCTAATGTAAATGACGAGTTAAT
GGGTGCAGCACACCAACATGGCACATGTATACATATGTAGCAAACCTGCACATTGTGCAC
ATGTACCCTAGAACTTAAAGTATAATAAAAATAAATAAATAAAATAAAATAAAATTGCTT
TGGCTTTCTGGACTCTTTTTTTTGGTTTTATATGAATTTTAGGATTTTTTTCTAATTCTA
TGAAAAATGGCATTGGTAATTTGATAGGGATTGTGTCGAATCAGTAGACTGCTTTAGACA
GCATGGTCATTTTAATAATATTGAATCTCTAATCCATGAGCCAGGGATATTTTTCCATTT
GTTTTGTCATCTAGGGTTTCTTCCATCAGTGTTTTGTAGTTCTCCTTATAGATATCTT
TTACCTCTTTGGTGAAATGTATTCCCAGGCATTTTACTTTATCTTATCTTTTTGTAGCTA
TTATAAATGGAATTGCTTTCTTAGTTTGGTCCTTGGAAATGCCAACTACATTTAAAATCC
TTTTCCATTTGATGGATTTCAGGTCTTGATGAACATCTCAGTTGTAATTTTCTTAAGATT
GAAAAAGTAAATATTTTTTCTATATGTATATATAAAATTGTCCTCTCTCAAAATTTTAAT
TCAATAACCTGCTAGATATCACTTTAGAATCTTGCAGTACTAGTTTTCTTCTCAATTAAT
TGTAGATCTTAGCCTTTTAATTTGGGCATGTTTTTCCCTATTAGGACTTAAGTTATTAGG
ACCTAAGTTTGTAGACAAGAACTATGTTATATTTGAGAAATTTGTGAGTCATGTACTGGG
CCTAGCACAGTGCCTCATAAGATGTAGACCCTCAATAAACTTGTTGAATAGGTTAATAAA
TAAAAAAGCCCCTATCACTCAATTTTTTTTTTTTTTTTTTAGATGGAGTCTCACTCAGT
CACCCAGCCTGGAGTGCAGTGGCACGATATCGGCTCACTGCAAGCTCTGCCTCCTGGGTT
CACACCACTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGACTACAGACACCCGCCACCATG
CCCGACTAATTTTTTGTATTTTTAGTAGAGACGGGGTTTCACCTGTGTTAGCCAGGATGG
TCTCGATCTCCTCACCTCATGATCTGACCCCCCTCGGCCTCCCAAAGTGCTGGGATTACAG
GCATGAGCCACCACACCTGGCCTATTTCAGTCAATTGTTAAAAGTGCTAAGAACAAGTGG
AGATCTTGTTAATGAAGAAAAAAAAAAATAGTATTTACTACTTACCTAAACACTCTACTAA
GAAGGGATATACAGATCAAAAGGATTAAATCTCTGCCTGCATTAAGCTAACTGTTTTGTA
AAGAAGAACGTAAACAAAGTCAAAAATGCATTTTTTAGGTGCTAGAGATTAGACAGGACA
AAATCTTCTGGCTCTGCCTAGAGTTAAGTGGCTTTGGGAGAGGCTTTGCTGTAGTTTAAA
GGCAGAGGTGGGGAAGGCCACTCTGGCCACAAGGACAGATCCACAATGGGATGGGGTATG
AAACAGCACGAACCCTTCAGGAAATTACACATAATTTAAAAGGAAAATGGGAGCCCATGG
CAGAAAATAGAATTGAACAGCAGGAAAAGGGTAGATAGTAAAAAGCATTTTATAATATTC
AAGGACATTTGAAACTTGTGGTATACAATGAGGAAGAATTTAAAAATTCTATACAGAGGA
GTGACATAGTTAGATTTGTGTTCTGGGGAGCATAATAATAGCATTACAGCGGGTGAATTT
GAAAGCTGGGCCTCAAAAGTTTAGATCTCAAATAGGTTTTTATGGGAGTATTCATCCTCA
TGAAACATGATTTGGAACTAAACCAAGGCAGTGGCAATGGGGCTGGAAAATAAACACTAG

FIG. 1AR

ATTTCATATCTAGATGAAGATTTGTGGAATAAGAGAGGCCACATTAATGTTTAATTCTAT
TTACAATGGATCCCAGCCACCATCCGCTTTAACACAGAGGTGCTTTTCCAGTAGCTAAGA
GGACTAGGTGCTTTAGATACATTTGTGAAGTTGTCCTCCCATTGTTAACATGCTTTTTTT
ATTGTCTGTGTGTAGGTTGATGGGGGAGGCAGAGTTAGGATCACACATAGAAGTTCAGTC
TTTGAAATGCTTTCTTTCTCTTTTTCCCCAAACAATGACCCCCACCTTTTCCTTCTGGCA
TATGTTGCCTCAAGACCCTAACACTGCTGCCAATCTGCTGGTCTTAGAGCCAAGAATCTG
CCACCACCTGGCCCACCACAGCCTGCTCTGCTAGCTGCTCTCCTGCCAATACTGGCCTTC
ATGTACAAGTGTAGGTTTTGAGGGTTCCGTTCTCTCCCCCTTTCTCTCTTTGAGTGTGGG
TTTGTGAGTGTGTGTGTCTTCTGTAATAAGAAGAAAACAGGCCACATTTCTCTACTCGT
GTTATACACTTCCCGGAGTGTCTCACATCAAAACCTGTCCTAAGTCCAAGCCTTAGAAGC
TCTTTGCTGGCCCAGCCTACACTTGGGTTGTTACTTCTCAGGAGCTACCTTTCTGTCAGT
TGAGATTTTAACAACCCACCACAGTACTCCAAGCGTGCAGTCCCTCACATCTTGAAATCT
GTGCTTTGGCAGCAGCAGAATCAGGGGTCTGTGGATTCTGAACCCAGAATGTGTCAAACC
AAAGGGTGACATATTGGGACATTTAATAAGTCAGAGACTATTTCCCAGGAATATTTTTTT
GAAGCATTTAAACTAAAATACAATTGAACTGAGATCTCAAAACAGGAAAAATGAACTTGA
CAAGAATTAGGCTAAGCTGCATCTCATGACGTAAATATTCACATTTGCATATACATTAAC
AGAGTCAAGTCAAAAATTGATTTTTTATTGGATAGGATTAACTTTAGCTACAGAAAACAG
AAAGTTTAAATGACTGGCTTTAAAAAGCAAAAGTTTACTTGCTTTCATTTATATGCAATC
TGGGGGGTTGCGGGGGAAGATTTGTTTGTGGGTTTCCCATTCTCAAGGTGCCAGGCTCCT
GTGTTTCTGCCCCATATCCTTAGAGGGAGGGTTTCCTCCTCAGGATTGCCTTATGTGCAA
GGTGACTACTGAAGCTCCATTTTTTATGCCCAAATTGTAGCAAGAAAGAGAAAAGGAAGA
AGGAACAGAAAGGCACATGACATCACTTCAAATAAAATTAGGGGGAAAATAATAACAGAT
ATCTGATAGGAAACTAACAGTACCTTCTGCAATAATGATTACATTTCTGGAAAATAAATG
TTAAGATCCTTGAAAACAAGGAGTAATAGTTGAGGAAAAGCTTCTTAGCAGCCTGGGACT
AAAAACTTCAAAAAATTTAAGATAAAAATCTGAAAACTGGTAGAGAATTGGGGAGAAAAA
GAGAATTTGAACAAGGCATGCAAGAGTAAGAAAAATGTCATCACAAAATTACTAAGAAAG
CATAAAAGCAACTATATTTTATTAGAGTAAAAATAAATGGATTGAATAACCCTAGTAAAA
TAAATTCCAGCCATACATTGTTTATAAAAAGTGTATTTAAGGTGATTAGGAAAAAATAAA
ACTAAATTTAAGTGCAAAGATCTCCCAGGGAAGTCAAAGCAAAAATAAAGTCAGATGTTG
CTGTCATTAGACAAAGTAAAATTTAAGGTGAAAACATGACAAAGAGGGACATTAAATAAG
GATAAATGTACAATCAATGGTGACAAACTTTTATAAACTTGAAATTATATTAACAAAACA
TAAATCATGTAAAACTAAAATACCTTGATAAAATGCAAATTATCAGGTAACAAGAATATA
TCTATTGCAGAAAGTAATATATCAAACTAAAATAATGTGTATCTATTACAAGTATACAAT
ACTTTGTAGCCTACAAAATAAGAATATACGATTTCTTCTTAAAATGTTATACATTTACAA
TAATTAATAATTTGGCCACTCAGAAAACCTTGGTAAAGCAAGGAAAGTAGAGATATTATA
AGCCAACTTAATAATTTAGTAACATTGGGTAAAAATGGAAGAAGTATCATATTGTGGTTG
TGAACATAAGCTCTAGTTCTCCTAGTTTTGTGATTTGGGAAAGTTAATTATCTTCTCTCT
ACCTCGTCTTAATTTTCAGTAATATTAGGATAACAATAGTTTGTACATCATCAGTGTTTT
TTTTTTTTGAGGAATAAATGACTCACATGTATTAAACACTTAGATCCATTGTTAACATAT
AATATGTATAAATAATGTCAGTATAAATCAATGTCAGCCTAAAAAGTTAAGACTGTGATT
TTAAATAATACTAGATTTAGAATAAAATCAAAATTGAAATGACATTATTAACTTAAAAAT
AACAAAAAAAGAGAAGACTTTAAACACAATGGATGGAAAGCAGCTATACCAATAAAAGAC
AAAAATGTGGAGTATTATATGTTCTTAATGTTTTTTAAAATTATAAAAATAAATAAACTA
AAACATAGAATTTTAAAAATTAAATGTTGGAGGGATTAGGTCAGATAAGAGAAATTTCTG
TTAGCAGCAGCTGAATTTTCTGCTAATAACAGAGAATTGTGAAAAGATGATTTCATAAAT
ATGGCAAATGTTTGTAATAGCCATCCTAGGAGCACGGATATTAGTAACTAATTGAGGAAG
TACTGTTGGGCAGTGTCAATATACTGGTTAAGAATAGAATTTAAAATAATGCTAATTATA
AGGCCAAAAAAACTCAGTAATGCAATTTTTTTAGTATAATTCAGTAGGGGAGAAGGAGAGA
TAATTAAACTTGGAAATTGACATACAGTTGTCCCTTGGCATCCATGAGGAATTAGTTCCA
GGACTCCCTATGGATACCTAAATTCACAAATGCTCAGGTCCCTTATATAAAATGGCAAAA
TATTTGCATATAACTTACACACACCCTCTTTATAATTTAAGTCATTTCTAAAGTACTTAT
AATACCGAATGCATTATAAATGACTGTGGAAATAGTTGTTGTATTATTTAGGGAATAATG
ACAATAAAAAATATATGTATATGTTCAGTAACAGATGCCTTTTTTAAAAAAAAATTGTTT
TTGATCCACAGTTGGCTGAATCTATGGATACAGAGCCCACATTTACTGAGGGCAGACTAT
ATTTAGAGTACTTAAGGATCACAAGGGACACACATCTGAGGGTACTGAAGAGTGGGAAGA
AATTACTAACCAGAGGGTCAGACTAGAAGGCAAGGAAGTGAAGCCAGGAGATGATTAGAA
AATAAGAAAATCATACAAGCCTGGAGATTATGTTGAAGTGTAAGAACATAATTAGAGTGA

FIG. 1AS

GAAACATGAGTCAAGGAAGAAGGAGATTGGTGCTTGAGAGATGTGGCAGACTGTATCTTT
CAAAGATGGCTACACCAATATATATCTCATTCCACAAGCTGTTTTTACCATGCTGTATTG
ACGCTCTTCCATATGGAGGTGGGGCCTATGTCCCCTCCCTTGAAACCAAATGAAACTTTG
TAATTGCCTTGATCAACAGATTGCAGTAGGAGTGATGCTGGATGATTTCAAAGGCTAATA
CACACAAGAAAATAATGGCTTTCATTTGACTCTTTCTTGGAACATGTGCCTTGGAAACCA
TGAGCTTATTTGCAAGAAGCTCAGCTATCCTAAAGTTTATCTACTGGGTAGACCAAGTGG
AGAAATTACACAGACATTGAGATTATGTTCAAGGGGTCTCAGAGGTTCAAGGCCTCCCAA
TTCAGGCACCAAACAAGTGGAGAAAAGGCTTTCAAGATCATCCCTCTGAAATAATTGTCT
GATTGAAACCTCAAAAGAGTCCCTGAGCCAGAACCATCCAGCCAAGCCACTCTCAAATTC
CAAATCCACAGACACCATGAATGACAGTAAATCATTATTGTTGTTTTAAAGCACATAAGT
TTTGGGGGGTTATTTACACACAGCAACAGAAAAAAAAACTGATGAATGGGAAACATGGAG
AGAAATGCAAATAGAATAAAATGGGAAGGAATACAAGGAGAGGAAAGTAGTATTGTGCAA
AATAGGCAATCGGATGACCCTCAAAAGGAAATTTTTTTCTGAGCAACTTAATGAATATA
AGGTCAGATTAAATTGGAAGGTAACAGGTACAAATATCATTAATGCTAAATTCTATTTGT
AGTAAGTCAACTATTTGTAAATTATGCATTGGAGACCGACTTTACATCAATCAAAAGTTA
AATTTATTTAGAAATCTATAGAAGAAGAAAAAGAATAAAAGCCATTGGAAAAGTTTTTAC
AATTATTCCATTAAATAGACAAAGTCCTTTAAGGAAAGGGATTAAAATGAAGGTAAGGTG
ATCTGCTTAAAAATAATATAGCAATCTGGGAGCCATGGCTCATGCCTGCAATCCCAGTGC
TTTGGGAAATCTAGGCAGGAGGACCTCCCAAAGGAGGACTTGGAGTTTGAGACCAGCCTA
GGCAACACAGAGAGACTCCATCTCAAAATTTTAAATTTCTTAAAAGAAAAAAAATAAAAT
GAAATAATATTGTATTAATTCCAGTAAAAGCATCAGACCAATTTAGAATATGGATGAGAG
AGAAAAACTAGAAATAACACCACAACAAGGAAGGAGAAAGCTGGTCTCTGGCAGGGACTT
CTAATTTAGAGAAAGACAGATGATAGCAAACAGCAAAAGTTGTATTATAGATGTAACTTA
AAAACTAATTTGATTTTTATTTTTAGTCAGAAAACTGCTTTAGGTATGGAACAAGTATAA
CCTGGTATTTCCAGTATCTCTCTGTTGACCTCACATCTCTCTCCAGATACTGCCTCAATT
CTCTGCTTCTCTTTATAGCAAATTCCCTTGAAAGAGAGACTACCTGGATCAGAAATTCCT
CTGCTTCAATTTGATCCTGAATCCACTTCATCTAGATCTTCCTCACCAATTCCCCCAAAT
ATTTGTCTTATTATGGTCACATGGGACCTCTACTTTGCTATATCAGTAATTTTGTTCTCA
TTTTACTTTTTTGTAGTTAATTACTCCCTTCTCCTTGAAACACTTTCCTTGTTTGGCTTC
TAGGATGCCCTGTTCTCATGGATTTCCTTTCACTTCTCCAGTCATTTCTGTTTGTTTTTT
CAATATCTTCGTGATCTTATATTTTTAATGCGTCCTGCTAGCTTCCCAACTAGGTTTCCT
ACTTTCACCTTAATTCCCTATGGTTTATTCTCTACAAGAAAGGAATTATAATCCCTTAAA
AATGTCAATAAAACTCTATCACTACTCAATACTCTCCAAGGGGTCCTTATTTTATTCAAG
TAAAAAACTAAAGTCCTTACTATATGTCTGTAAATTCCCATAGGATCTGGCCCCACAGCC
CCTCTGGCCCCTGTCCATTCTGCCCCTTGCCAATTCTGCCCGGCCACAGTTGCCCAATAG
CTGGTCTGTGAACACATCAAGCACATACTTAATCTCAAGGCTTTTGCAATCATTCTTTTC
TCTAGTTGTAATCTCTCATTACTTATTCTGAGTGTCTTGTTTCTGCAGTTGCTTTACTTA
CTTGACCTATATAAAATAGTAATTCTTACCCCTACAACTCCATTATGTCCTATCTTCTTT
GCCTTGCCTTATGTTTTTTTCTTGGAGTTACAGATACCTGATGTAGATAGTATTTACTTT
TTTTATGCTTGCATTAATCACCTAGAATATAAACTCCAAAAGAGGAGCTATTTCTCTTTT
ATAATCTATCTAATATATCTTGGATATTTGCTCCCACCTAAATTTCATGTTGAAATGTAA
TTCCCTGTGTTGGAGATGGGGTCTGGTGGGAGGTATGTGGATCATGGGGCGGATCCCTCA
TGAAGGGCTTGGGCCATCCTTTTGGAGAGAAGTGGGCTCTGGCTCTGACTTCACACGAGA
TCTGGTTGTTTAAAAGTGTGCGACAGCTCCCCTGAGCTTCCTCTCTCACTTGCTCCTGCT
TTTGCCATGTGAAGTACCAGCTACTGCTTCATTTTCCACCATGAGTAAAAGATCCCTGAG
GCCCTCCCCTCAGCAGTACATGTCCCTATGCTTGTTGTGCAGCTGGCAGAACCATGAGCCA
ATTAAATCTCTTTTCTTTTAAATTACTCAGTCTCATGTATTTCTTTATAGCAATACAAGG
TTGGCTTAATACATATCTCTAAAGCAAAAGCTGGGCCTGGTATGTAATAGGTGTTCAATA
AATATTTATTGAATAAATGAATAAATACTAGGCTAAATAAAGTTTAAAACATCATAATAG
AACACTGGGTAGATGTCAAGATGACAGTTTTGTTATTCACATATGGACATGGAAAGGTCT
TTGTGGTGCATTGTTAAGGGAGCAAACCAAATTACAGAACACTATATAGAGTAGAGCTGT
ATAAAATACATATGGTGTATGTTTATAAATATGTCTAGAAAAATTTGAAAGCTATATATC
AAATATCATATCATTTATCTTTAGAAGGCTAATTGCATATTTTCAATTTATTGTTTATAA
TTTTTTTTATCTATTATTATAGGTTACTTGTATAATCACAAAAGACAACTGAATAATTCT
TTTTGTCTTCATCAACTTTTATTTTAAGTTCTGGGATACATGTACAGGATGTGCAGGTTT
GTTACATAGGTAAACGAGTGGCATGGTGGTTTGCTGCACAGATCGACCCATCACCGAGGT
ATTAAGCTCAGCATCCATTAGTTATTCCTCCTGATGCTCTCCTTCCCCTTGGCCCACCAA

FIG. 1AT

TACACCCTAGTGTATGTTGTTACCCCTCATGTGACCATGTGTTCTCATCATTCAGCTCCC
CCATATAAGTAAGAATATGCAGTGTTAGGTTTTCTGTTCCTGTGTTAGTTTGCTGAGGAT
AACAGGTTCTAGATCCATCCATGTCCCTGCAAAGGACATGCTCTTGTTCCTTTTTATGGG
TGCATAGTATTCCTTGGTGTATATGTACCACATGTACAACTAATTTCCACAACAAAAAAT
GTACTATTACATGGATATAATGTTTATATTCTCTTCACAGAATTTGAGTCACTTGAATTT
TTGCTTTAACACTTAGAATTTGGAGGGTCTGTTTTCTTAAAAAAAATTAACACTTTAAAT
CCAATAAGTAAATGTGGAAGGTTGGTGGAAATAGTTAGCTGGAAACTCAGAATTGATATT
AACTTTCACCAAGCCTTTGTTCACATTATTTTCTTCTACAATTTATGAATGAATAATCCT
GCACTATCTATGCATTCAAACAATGATACATATGGTGCATATGTATATATGGCAAAAATC
TAAGAAATGTAGCCAAATATTAATATTGCTTACACGTAAGTAGTCAAATCATGGTGGTTT
TTTTTATTTTCTTGATTTTGCAAGAAAATTAATAAAGAGGCTATTTACATTTTAATGTA
CAAATGTGTATACAAATATAATAGTTATGCTTTAAAAATCCAATAAATAAATGTAAGTAA
AACATTTCTGAATTTTTTAAAGATTTCTCAATAGATCTAGGTATTCTTCTTAACCAAATA
CTGATACTACCGTTAACCACTTCTGGAAAATTCTGGCAATTGGTCCCTTTGGGGAAGAAC
TAGAGGAATCACTACTATACACACTTACTGTGGTATTCAGTGCCCTTCCTCAAGGGGAAT
TCGCCTATCTTTTTTTTCTTAAGTAATATTTTATCTTTAATAGACAAATAATGGTTGTAT
TTATTTACGGGATACAAAGTGACATTTTGATGCAAGCATACCTTGTGGAATGATCAAATC
AGGCTAATTAACATATCTGTCATCTCAAATGCTTATCCTTTCTTCATTGTGGGAGCACTT
AAAATCAATTCTTTTAGCTATTTGGAAATATAAAATATATTATTTTCTAACTATATTTAC
TTACGATGTAGTGTAATAGATCACAAGAACCTATTTCTTCTATCTAACTGAAACTTTGTA
CTCTTTGACCAACATCTCCCCTTTCTTTGTCCATCCTCCTAGCCCAGCCTTTGGTAGCCA
TCACTGTACTCTGTATTTCTATCACTTTGCCTTTTTAAATTGCACATATAAGAGAGATCA
TGCAGTATTTGTTGCTTTGTGTCTGACTTATTTCCTGTAGCAGAATGTCCTTTAGGTTAA
TCCATGTTGTCATAAATGACAAAATTTCCTGCCTTTCAAAGGCTGAATAGTATTCCATTG
TTTATATATACCACATTGTCAAAATCCATTCATCTGTTGATGGGCATGTAAGTTGTTTTC
AAATATTGGCTTTATTAATAATGCGGCAGTGAACGTGGGAGTTCAGACATCTTGTTGACA
TACTGATATTAATTCCTTTGACTATATACTCAAAAGTGGAATTGCTGGACTGTGTGGTAA
TTTTAGATTTTTAGTAACATTCATACTGTTTTCCAAAATAACTGTATGAATTAACAATAC
CATCAACAATGTACAAGGGTTCCCTCTGCTCCACATCCTCATCAACACTTGCTAGTTTTC
ATGTTTTCGATAATAGCCAGTCTATCAGGTGTAAGATAATATTTCATTGTGATTTAATTA
GCATTTCTTTGATAATCAGAGATTTTGAGCCTTTTTTAATATATCTGTTGACCACTTTTA
TGTTTTCCTTTGAGAAATGTGTATTTAAGTCGTCTGCCCATTTTTAATAGGATCATTTGT
TTTCTTATTATTGAGGGGTTTGAGTTCCATGCATATTTTAGATACTAGCCTTTTATCCAA
TGCGTAATTTGCAAATATTTTCTCCCAATCTGTGGGTTGTCTCTTTAACCTGCTAACTGT
TTCCTTTCCTTCCTGCAGAAGCTTTTTAGTTTGATGCAATTCCATTTGTCTATTTTTGCT
TCCATTGCCTGTGCTTTTGGGGTTAAGAAATCTCTGCTCGATTACATTTATTGATTTGCG
TATATTGAACCAGCCTTGCGTCCCACGGATGAAGCCCACTTGATCATGGTGGATAAGCTT
TTTGATGTGCTGCTGGATTCGGTTTGCCAGTATTTTATTGAGGATTTTTGCATCAATGTT
CATCAAGGATATTGGTCTAAAATTCTCTCTTTTGGTTGTGTCTCTGCCAGGCTTTGGTAT
CAGGATGATGCTGGCCTCATAAAATGAGTTAGGGAGGATTCCCTCTTTTTCTATTGATTG
GAATAGTTTCAGAAGGAATGGTACCATTCCTCCTTGTACCTCTGGTAGAATTCGGCTGTG
AATCCATCTGGTCCTGGACTCTTTTTGGTTGGTAAACTATTGATTATTGCCACAATTTCA
GAGCCTGTTATTGGTCTATTCAGAGATTCAACTTCTTCCTGGTTTAGTCTTGGGAGGGTG
TATGTGTCAAGGAATTTATCCATTTCTTCTAGATTTTCTAGTTTATTTGCGTAGAGGTGT
TTGTAGTATTCTCTGATGGTAGTTTGTATTTCTGTGGGATTGGTGGTGATATCCCCTTTA
TCATTTTTTATTGTGTCTATTTGATTCTTCTCTCTTTTTCTCTTTATTAGTCTTGCTAGC
GGTCTATCAATTTTGTTGATCCTTTCAAAAAACCAGCTCCTGAATTCATCCATTTTTTGA
AGGGTTTTTTGTGTCTCTATTTCCTTCAGTTCTGCTCTGATTTTAGTTATTTCTTGCCTT
CTGCTAGCTTTTGAATGTGTTTGCTCTTGCTTTTCTAGTTCTTTTAATTGTGATGTTAGG
GTGTCAGTTTTGGATCTTTCCTGCTTTCTCTTGTGGGCATTTAGTGCTATAAATTTCCCT
CTACACACTGCTTTGAATGTGTCCCAGAGATTCTGGTATGTTGTGTCTTTTTTCTCGTTG
GTTTCAAAGAACATCTTTATTTCTGCCTTCATTTTGTTATGTACCCAGTAGTCATTCAGG
AGCAGGTTGTTCAGTTTCCATGTAGTTGAGCAGTTTTGAGTGAGTTTCTTAATCCTGAGT
TCTAGTTTGATTGCACCGTGGTCTGAGAGACAGTTTGTTATAATATCTGATCTTATACAT
TTGCTGAGGAGAGCTTTACTTCCAACTATGTGGTCAATTTTGGAATAGGTGTGGTGTGGT
GCTGAGAAGAATGTATATTCTGTTGATTTCGGGTGGAGAGTTCTGTAGATGTCTATTAGG
TCTGCTTGGTGCAGAGCTGAGTTCAATTCCTGGATATCCTTGTTAACTTTCTGTCTCGTT

FIG. 1AU

GATCTGTCTTATGTTGACAGTGGGGTGTTAAAGTCTCCCATTATTATTGTGTGGTAGTCT
AAGTCTCTTTGTAGGTCACTCAGGACTTGCTTTATGAATCTGGGTGCTCCTATATTGGGT
GCATATATATTTAGGATAGTTAGCTCTTCTTGTTCAATTGATCCCTTTACCATTATGTAA
TGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGTCTGTTTTATCAGAGACTAGG
ATTGCAACCCCTGCCTTTTTTTGTTTTCCATTTGCTTGGTAGATCTTCCTCCATCCTTTT
ACTTTGAGCCTATGTGTGTCTCTGCACGTGAGATGGGTCTCCTGAATACAGCACACTGAT
GGGTCTTGACTCTTTATCCAATTTGCCAGTCTGTGTCTTTTAATTGGAGCATTTAGTCCC
TTTACATTTAAAGTTAATATTGTTATGTGTGAATTTGATCCTGTCATTGTAATGTTAGCT
GGTTATTTTGTTTGTTAGTTGATGCAGTGTCTTCCTAGCCTCTATGGTCTTTACAATTTG
GCATGATTTTGCAGTGGCTGGTACTGGTTGTTCCTTTCCATGTTTAGTGCTTCCTTCAGG
AGCTCTTTTAGGGCAGGCCTAGTGGTGACAAAATTTCTCAGCATTTGCTTGTCTGTAAAG
GATTTTATTTCTCCTTCACTTATGAAGCTTAGTTTGGCTGGATATGAAATTCTGGGTTGA
AAATTCTTTTCTTTAAGAATGTTGAATATTGGCCCCCACTCTCTTCTGACTTGTAGAGTT
TCTGCCGAGAGATCCGCTGTTAGTCTGATGGGCTTCCCTTTGTGGGTAACCCGACCTTTC
TCTCTGGCTGCCCTTAACATTTTTTCCTTCATTTCAACTTTGGTGAATCTGACAGTTATG
TGTCTTGGAGTTGCTCTTCTCGAGGAGTATCTTTGTGGCATTCTCTGTATTTCCTGAATC
TGAATGTTGGCCTTCCTTGCTAGATTGGGGAAGTTCTCCTGGATAATATCCTGGAGAGTG
TTTTCCAACTTGCTTCCATTCTCCCCGTCACTTTCAGATACACCAATCAGACGTAGATTT
GGTCTTTTCACATAGTCCCATATTTCTTGGAGGCTTTGTCCGTTTCTTTTTATTCTTTTT
TCTCTAAACTTCCCTTCTCACTTCATTTCATTCATTTCATCTTCCGTTACTGATATCCTT
TCTTCCAGTTGATCGCATCGGCTCATGAGGCTTCTGCATTCTTCACGTAGTTCTCGAGCC
TTGGCTTTCAGCTCCATCAGCTCCTTTAAGCACTTCTCTGTATTGGTTATTCTAGTTTTA
CATTTGTCTAAATTTTTTTCAAAGTTTTCAACTTCTTTGCCTTTGGTTTGAATTTCCTCC
TGTAGCTCGGAGTAGTTTTATCGTCTGAAGCCTTCTTCTCTCAACTTGTCAAAGTCATTC
TCCATTCAGCTTTGTTCCATTGCTGGTGAGGAGCTGCGTTCCTTTGGAGGAGGAGAGGTG
CTCTGCTTTTTAGAGTTTCCAGTTTTTCTGCTCTGTTTTTTCCCCATCTTTGTGGTTTTA
TCTACTTTTGGTCTTTGATGATGGTGATGTACAGATGGGGTTTTGGTGTGGATGTCCTTC
CTGTTTGTTAGTTTTCCTTCTAATAGACAGGACCCTCAGCTGCAGGTCTGTTGGAGTTTG
CTAGAGGTCCACTCCAGACCCTGTTTGCCTGGGTACCAGCAGCGGTGGCTGCAGAAGAGC
GGATTTTCGTGAACCGCGAATGCTGCTGTCTGATCGTTCCTCTGGAAGTTTTGTCTCAGA
GGAGTATCCTGCCGTGTGATGTGTCAGTGTGCCCCTACTGGGGGGTGCCTCCCAGTTAGG
CTGCTCGGGGGTCAGGGGTCAGGGACCCACTTGAGGAGGCAGTTTGCCCGTTCTCAGATC
TCCAGCTGCGTGCTGGGAGAACCACTGCTCTCTTCAAAGCTGTCGGACAGGGACATTTAA
GTCTGCAGAGGTTACTGCTGTCTTTTTGTTTGTCTGTGCCCTGCCCCCAGAGGTAGAGCC
CACAGAGGCAGGCAGGCCTCCTTGAGCTGTGGTGGGCTCCACCCAGTTCGAGCTTCATGG
CTGCTTTGTTTACCTAAGCAAGTTTGGGCAATGGCGGGCACCTCTCCCCCAGCCTTGCTG
CCACCTTGCAGTTTGATCTCAGACTGCTGTGCTAGCAATCAGCAAGACTCTGTGGGCATA
GGCCTCTCCAGCATATAAACAGAACCAAAGACAAAAACCATATGATTATCTCAATAGATG
CAGAAAGGGCCTTTGACAAGATTCAACAACGCTTCATGCTAAAAACTCTCAATAAATTAG
GTATTGATGGGATGTATCTCAAAATAATAACAGCTACTTATGACAAACCCACAGCCAACA
TCATACTGAATAGGCAAAAACTGGAAGCATTCCCTTTGGAAACTGGCACAAGACAGGGAT
GCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGAAGTTCTGGCCCAGGCAATTAGGC
AGGAGAAGGAAATAAAGGGTATTCGATTAGGAAAAGAGGAAGTCAAATTGTCCCTGTTTG
CAGATGACATGGTTGTATATCTAGAAAGCCCCATTATCTCAGTCCAAAATCTCCTTAAGC
TGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAATCAATGTACAAAAATCACAAGAAT
TATTACACACCAATAACAGACAAATAGAGAGCCAAATCATGAGTGAACTCTCATTCACAA
TTGCTTCAAAGAGAATAAAATACCTAGGAATCCAACTTACAAGGGACGTGAAGGACCTCT
TCAAGGGAAACTACAAACCACTGCTCAATGAAATAAAAGAGGATACAAACAAATGGAAGA
ACATTCCATGCTCATGGTTAGGAAGAATCAATATCGTGAAAATGGTCATACTGCCCAATG
TAATTTATATATTCAATGCCATCCCCATCAAGCTACCAATGACTTTCTTCACAGAATTGG
AAAAAACTACTTTAAAGTTCATATGGCACCAAAAAAGAGCCCGCATCACCAAGTCAATCC
TAAGCCAAAAGAACAAAGCTGGAGGCATCACACTACCTGACTTCAAACTATACTACAAGG
CTACAGTAACCAAAACAGCATGGTACTGGTACCAAAACAGAGATATAGCTCAATGGAACA
GAACAGAGCCCTCAGAAATAATGCTGCATATCTACAACTATCTGATCTTTGACAAACCTG
AGAAAAACAAGCAATGGGGAAAGGATTCCCTATTTAATAAATGGTGCTGGGAAAACTGGT
TAGCTATATGTAGAAAGCTGAAACTGGATCCCTTCCTTACAGCTTATTCTAAAATTAACT
CAAGATGGATTAAAGACTTAAACGTTAGACCTAAACCATAAAAACCCTAGAAGAAAACCT

FIG. 1AV

AGGCATTACCATTCAGGACATAGACATGTGCAAGGACTTCATGTCTAAAGCACCAAAAGC
AATGGCAACAAAAGCCAAAATTGACAAATGGGATCTAATTAAACTAAAGAGCTTCTGCAC
AGCCAAAGAAACTACCATCAGAGTGAGCAGGCAACCTACAAAGTGGGAGAAAATTTTCGC
AACCTACTTATCTGACAAAGGGCTAATATCCAGAATCTACAATGAACTAAAGCAAATTTA
CAAGAAAAAAACAAACAACCCCATCAAAAAGTGGGTGAAGGATATAAACAGACACTTCTC
AAAAGAAGACATTTGTGCAGCCAAAAAACACATGAAAAAATGCTCATCATCACTGGCCAT
CAGAGAAATGCAAATCAAAACCACAATAAGATACCATCTCACACCACTTAGAATGGCAAT
CATTAAAAAGTCAGGAAACAACAGGTGCTGGAGAAGATGTGGAGAAATAGAAACACTTTT
ACACTGTTGGTGGGACTGTAAACTAGTTCAACCATTGTGGAAGTCAGTGTGGCGATTCCT
CAGGGATCTAGAACTAGAAATACCATTTGACCCAGCCATCCCATTACTGGGTATATACCC
AAAGGACTATAAATCATGCTGCTATAAAGACACATGCACACGTATGTTTATTGTGGCACT
ATTCACAATAGCAAAGACTTGGAACCAACCCAAATGTCCAACAATGATAGACTGGATTAA
GAAAATGTGGCACATATACACCATGGAATACTATGCAGCCATAAAAAAGGATGAGTTCAT
GTCCTTTGTAGGGACATGGATGAAATTGGAAATCATCATTCTCAGTAAACTATTGTAAGA
ACAAAAAACCAAACACCGCATATGCTCACTCATAGGTGGGAATTGAACAATGAGAACACA
TGGACACAGGAAGGGGAACATCACACTCTGGGGACTGTTGGGTGGGGGGAGGGGGGAGGG
ATAGCCTTAGGAAATATACCTAATTATAAATGACGAGTTAATGGGTGCAGCACAGCAGCA
TGGCACATGTATGCATATGTAACTAACCTGCACATTGTGCACATGTACCCTAAAACTTAA
AGTATAATAATAATAAAATAAAAAAATAAAGAATAGAATAAATAAAAACAAAAATAAATA
AAAAATAAAAAGAAATCTCTGCTCATATCCAGGCCATGATGGTTTCCCCCTGTGTTTTCT
TCAAGTAGTTTTATAGCTTCAAGTCTTATGTTATATTAAGTCTTTAATCCATTTTGAGGT
GATTCTTGTACAAAGGCTGAAGTAAGGGTTCATTTTGATTCTTCTGTGTGTGTGTATCCA
GTTTTCCCAACACCATTTATTGAGAAGTCTGTCATTTCCCCATGGTGTGATCTTGTTACC
TTTATGAAAATTTAATTGACCATAGGTGTATGGGTTTATTTCTGGGCTTTCTATCATATT
CCATTGATTGATATGTCTGGTTTTATGCCAGTACTATGCTGCTTTGATTACTGTGGATTT
GTAATGTAATTTAATGTCTGAGAGTGTGAAGCCTGCAGCATTATTTTTTCTCAAGATTGT
TATCTGTGGCTATTTGTAGTCTTTTGTGGTTTCATATATATTTTACAATTTTTTATTTCT
GTGAAAAATGCATTGGAATTTTCATATGGATTACATTTAATCCGCTTTGGGTAGTATGAC
CATTTTAACAATATTAATTGTTCTAATCCATGAGCATGGGCTAGCTTTTCATTTATTTGT
GTCATCTTCAGGTTTTTTCAACAATGTTTTATAGTTTTAGTATATGGATCTTTCACTTCC
TTGGTTAAATTTAGTCCTAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTTTGTGTGTAT
GTGTGTGTGCATCAACTAACCATAGTCATGTGGGTTTATTTCTGGGCTTTCTATCATGTT
CCATTGATTACTTCTAAGTGAATGAGTGTGTGTGTGTGTGTGTGTGTGTGTTTAAGAT
ACTGTTGTAATTTTAAAATTTCTTTCTCAGGTTGTATGTTGTTAGTGTACAGAAATAATA
TTAATTTTGTAAGTTGATTTTGTATTCTGCAAATTCACTAAATTTGTTAATTTGTTTTAA
CAATTTTTTGGGTGTAGTCTTACAGGGTTTTCTATATATAAGATCATGTCATCAGTAAAC
AATTTCATTTATTCTTTTCCTATTTGGATGCTTTTTATTCTTACCCAATTGTTTTGACTA
GGACCTCCAGTACTATGTTGAACATAATTGATGAAAGCAGACATCCTTGTCTTGCTCCTG
ATCCAAAAGCCTTTAACTTTTCACCACTGAGTATGATGTTCACTGTAGGCTTGTTATATA
TGGTCTTTGTTGTGCTGAGAAACATTCCTTCTATAACTGATTTTCAAAAGTTTATCATGA
AAGGATGTTAAATTATTTCAAATGTTTTTTCTTCATCTATTGAGGTGATTATATTGTTTT
TATTCTTCATTCTGTTACTATGGTGAATCATATTTTTAATTGTTTTTTACTTGCATAAAT
TTATTTTGTGATAGGTAGAAAAGCACATCTGCAGACCTAGAAGCAGAGTGAATCTAAAAA
ATATTATTTATAATTATTATGAGTACACAATAGGTATATATTTTCATGGGGTACATTCAA
TGTTCTGATACAGGCATATGATGTGTAATAATCACATCAGGGTATTTGGAGTATTCATTA
CCTCAAGCATTTATCATTTCTTTGTGTTAGGGAATTTCAGTTTCATTCTTCTAGTTATTT
AAAATATACAATGAATTATTATTGACTGTAGTCACCCTGTTGTGCTATCAAATAGTATGT
CTTATTCATTTTATTTAACTATATTTTTGCACCCATTAACAATCCCCACTTGATTTGAAT
ATGGTAAGCCATTCTTGCATCCTAGGAATAAATTCCATTTGACCATGGTGAATGATCCTT
TTAATGTACTGTTGAATATAGTTTTTGGTATTTTGTTGAGGATTTTTGCATCCATGTTCA
TCAGCGATATTGGCCTGTAATTTGCTTTTCCGGTAGTTTCTTGTTTTTTTATTATACTTT
AAGTTTTAGGGTACATGTGCACAACGTGCAGGTTAGTTACATATATATACATGTGCCATA
TTGGTGTGCTGCACCCATTAACTCATCATTTAACATTATGGAAAATCTCCTAATGCTATC
CCTCCCCGCTCCCCCCACCCCACAACAGGCCCCGGTGTGTGATGTTCGCCTTCCTGTGTC
CATGTGTTCTCATTGTTCAATTCCCACCTATGAGTGAAAACACACGGTGTTTCTTAGTCT
GGCTTTGGTCTCAGGCTAATGTTGGCCTTACAAAATGATTGTGGAAATATTTCCTTCTCT
TCAATTTTTTGAAGAAGTTTGAAAATAATTATTACCAGTTCTTCTATAAATGTTGGGTAG

FIG. 1AW

AATTCATTTATGAAAATATCTTTTCCTGGGTTTTCCTTGATGGCGGACTTTTCATTACTG
ATTTAATTTCCTTGCTCATTACTGTTCCATTTATATTCCTCATGATTTGATCTTGGAAGG
TTATGTATCGAAGCCTTTATCTATTTCCTCTCCATCGTCCAATTTGTTTGCATGCAATTG
TTCGTAGTGGTCTCATAAGATCCTTTGTATTTTTGTACTATCAATTGTGATATCTTTTTT
CATTTCTGCTTTAGTTTACTTGAACCACCTGTATTTTCTCGTGGTTAATTTAGCTAAGGA
TTGTCAATTTTGTTTGTCTTTTTGGAAGACCAACGCTTAGCTTTACTGATCTCTTGTATT
GTTTTTCTAATTTCTATTTCATTGATTTTTGCTCTGAAATGTTTCCTTTCTTCCACTAAC
TTTAGGCTTAGATTGTTCTTCTTTTACTAATTCATTGAGGAGTAACATTAAGTTGTTTAT
TTAAGATCTCTCTCTCCTTCTCTCACTCTCTCTTTTGATGTAGGCATTTAGTGTTACAAA
CTTTCCTCTTAGAACTGCTTTTGCTGAATCCTGTAAGTTTTAATATGTTGTTTCCATTTT
CATTTTTCTCTAAATATTTTTAAAATTAATTTTTGAATTTCCTCTTTGACTCAATAGTTT
TTCAGGAGCATGTTGTTTAATTTGCATATACTTGTTAATTTTTCTTGGTTTCTCCTGTTA
TTGATCTATAGCTTTATATCATTGTGATTGAGAAAGATACTTGATATAATGTTGATCTTC
TGACACTTGTTAAGATGTTTTGTGGTCTATCAATTGATTTATCCTAGTGAATGTTACATG
TATACTTGAGAAAAATGTATATTTTGTTGCTGTTGGATGAAATGTTCTGTATAGGTCTAT
TAACTCCATTGGTATACGTATAGTTCAAGTCATATTTTGTTATTAAAAATTTTTTGTCTA
GATAATAGTTCTGTTGTTGGAAGTGGGATATTAAAATTATTTACTATTATTGTGCTGCAT
TTATGTCTCTTTTCAGAACTCTTAATCTTTGATTTATATATTTAGGTGCTTCAGTGTTGG
GTGCATATATATTTACAATTGTTATATTATCTTGATGCACTGATCTTTTTATTATAATAT
ACTGACCTTCTTTATCTCTTTTTACAGTTTTTTTTAACCTAAAGTTTATTTGGTGTGAAA
TAAGTATAGCCACCCCTGCTCTGTTTTATTTGCCTGGAATATCATTTTCCATCACTTCAT
TTTCAACCTGTAAGTTTCCTTTAAGGTAAGGTGAGTCTTCTGTAGGCCCATATAGTTGGA
TCTTGTTTGGTATGTATCATGGTACTGTATGCCTTTTGACTACAGAATCTAATCCATTAA
ACTTTAAAGTAATTATTGATAGATGAGAGGTTGCTACTTCCATTTTATTGTTTTCAAGTT
GTTTTCTAGATCCTACATTTTTTTTCTTATATCTTGCTTTCTTTACTTGTGATTTGATTG
CTTTTTGCAGGGATATATTTTGAATTTTTTAAAATATTTTGTGTATCTATTATAGGCTCA
TGCTTTGTGGTTACATAAATCATCTTATACCTATAACAAGCTATGCCAAGTTGATAACAA
CTTAAGTTTGATCACTTACACAAAGGCTACACTTTTACTCTCCTCCTTCTAAATTTTATG
TTTTTGATGTCATTCTTTACATCTTTTTATAATATGCATACTTAACAAACTACTGTAGCT
GTAGTTGCTTTTAAGAATTTTGCCTTTTAACCCTTATACTAGAGAAATCCTTGATTTGTT
CACCATCATTACAATATTAGAATGTTTTGGAATTGAAAAAATGCCATTAATTTTACCAGTG
CGTTTTATACTTTCATATGTTTTCATGTTTCTATTTTGAATCCTTTTCCTTCAGCTTGAA
GAACTCCCTTTAGCATTTCTTATAACGCAGGTCTAATGGTGAGAAACTCAGCCTTTGTTA
CTCTGAGAAAGTCTTTAACATCCCTCATTATTTAAAGACAGGTTTGCTAGGTATACTATT
CTTGATTGGCAGGTTTTTTTCTTTTAGAATTTTGAATATATTATCCCACTCCCTTGAGCT
TTCAAGGTTAATGCTGAGAAATTTGCTGATAGTTTTATCAGGGTTCTCTTATATGTGACA
ATTCAATTCTCCCTTGCTGCTTTCCATACTCTAAGTTTTGACAGTTTTGTTATGATGTGC
CTTGGTGTGAGTTTCTTTTCCTTTTTTAAATTTTAGATTCAGAGGGTACATGTGCAGATT
TGCTGCAAGGACATATTGTGTGGCGTTGGGCTTCTGTTGATCCCACCACTCAAGTGGTGA
ACATAGTATCTAGTAGGAAGTTTTTTGTTTTTTTGTTTTTTTTTAGCTCTTAGACCCTT
CTTTTTCCCTTTTTGGAAGATGCAGTGTCAATTGTTTCTATATTTATGTCTGTGTGTACC
CAATATTTAGTTCCTACTTATGTGAAAGAACATGCAATATTTGGTTTTCTGTTTCCGTGT
TAATTTGCATAGGATAATATTTTCCAGTAGTCTGTCCATGTTGCTGAAAAAGACATGAGT
TTGTTCTTTTTTATGGCTTCACAGTATTTCATGATGTATATGTACTTGGTGTGGATTTAT
CCGGATTCATTTTATTTGGTATTCTTTGGGATTCCTGTATCTGGCTTTCTATTTTCTTCC
CCAGTACTGGGAAATTTTCTGCCATTATTTTTTGAATATGTTCTGTGCTTGTCTCTCTCT
CCTCCTTCTGAACACCTATAATGTATATATTGCTCTGATTGAGGGTGTCAGTATGTCTCT
TAAGATGTGTTCATTCTTTTTCATTCTTTTTCCTTTTTGCTGCTTAGATTGGATGATTTC
CAGTGACTTGTCTTTGAGTTCATTGATATTTTCTTCTGCTTAATCTCATTTGTGGGTGAA
CCTTTCTGTCAATTTTTTCAGTTTAGTTTAATATTCCTCAGCTCTAAGATTTGATTGATA
CTTTCATATACTTTCTCTTTGTTAAAGTTCTCTGTTTTTGCATTTCTCTCTGGACCTTAG
TGACAGTCTTTATAATCATTATTTAAATTCTCTATTGGGTAAATTACATCTCTTCTATT
CACTTGGGTCAATTTCTGAACATTTATTTTGCTCTTTATTTGGAATATATATTTCTTGTT
TCTTTAGTTTCCTTGACTCTGTGTTGTTTACTGCACATTAGATAAGACAGCTGCCTTTCC
CAGTCTTATCAAACAGGACCTGTGTAGAAGAAAAAATATCACTAGTCCATTTGACAAAAAA
TTTTAATGTGCCTCTCAAAGCTTTGTTTGTCCAGGCCACTGTTTCTGTTATTGGTGGCTC
CCAGGAGATTGGGATATGCCATGTCCTATCAATACTCTGTGAACTATAAGATAGAGGCCA

FIG. 1AX

GACTTTCAAAATGTAGCCAGAAAAATGTCAAGTATTAGATGTGTGGTCCAGTTCCTTCTA
TCCTCATGTTGAAATTGGGTGCAGGTGTTACTTCTCCACTCTCTCTGCATGAAGCCAGGG
AGAGGTACTATGGAAACTGCCTGTATTTGTGTTCAGGCCACACTTTTTGATTCTGGGAAG
ATAGCTTTGGGAGTGGGGCCACTGTTTGTCTACATCTTTGTTATCTGTGATCTAGAGTAA
GTTAGGAATGCAAAGCTCCACCACTCCCAAGCTTAGGCTGTTAAGAATTCAGTCCTTTGG
GTGGGAGCTGTAGAAGTTGTGACACTTAATTGTGAACAAACTCTTTTCAAGAAGAATAGG
TAGGCTATAAAATAATAGAAGAAATGAATAGAGCTATAGAAGTTGTGACACTTGGTATGT
GAACAAACTCCTTTTAGGAAAAATAGGCTGGGGACAAGCCAAGTTCTGCTTAGTCTACCT
GAGAGCTACTATTAGTCTGTCTTGTTAGCTCCCTGATGCAAGCTGGAGGTTAAGCTATGT
AGTTGTCACTGGATGAGTGTGCAGTAAGCTGCTAGAGAAAAAAAAAAAAAAGGAGCTGTGC
ATTCTAGCCCCTGTTCTCCACTGCTCCCAAGAGATATAGTTCCTGGAAGAGTTTGCATGC
CTGTTTAAAACCACCTCTTTGTTCTGTGATCTAGGGAGACTTGTATATGCCTAGTCTCTT
CTGCTCTTAGAGCCAGGAGTTTTGGGATATAGTATTTCTGGTAAATGCTGTAAAAGGGCA
TTTTGTGGGTGAACACACTCCTTCCAGGGAGAATTGGGAGAGCTGGGATTATTGCTGAGT
TGAGCTGGAGGAAGTCTCAGGAAGTGTTAAGCTGCTGCTCAGGCTGTTAGAGAGCTACTT
TTTGCTTGCCCCTTTAACTCTCAGATGCGTTAGTTAGAAACCAGACTGTCAAGTAGCCGC
TAGGGGAGTATGCTGTAAACCTCTTCCAGGGAGAACCAGGTAGTGGTATTTTTGAGTCCT
GTCTCTGTACTAATTCTACTAATTCACAGTGTTAAAGCACCTGAAAAAGTGCTTGCACAC
ACATATAAAACTGCCACTGTTTTCCTGTGGTCTAAATAAACTTGTGTATGTCTAATTCTC
TCTAACTCCCAGAGTTGGTGAATTAAGAGCCAAACTGTTGGGCATCTTATAATTGGGGTG
CCATATGTAAGGTCCCAATCCTCTCCACAGGGAGAATCTGAGTGTTAGTGATTCCAGTTA
TATGGTGAAGTACCTGGAAGGGGTCCATGCTCAAGTATGCCTCAGATTTGTCTACCCATT
TGAAGTGCATGTTTGGGTTTTTATTTTGCTTTTGATGTGTTTTTTTTTGTCTTTTTTTTT
GAGACAGAGTCTCATTCTGTTGTAAAGGCTAGAGTGCAGTGGCACAATCTTGGCTCACTG
CAGCCTCTGCCTCCCTGGTTCAAGTGATTCTCCTGCCTCAGCCTCCCGAGTAGCTGTGAC
TACAGATGCGTACCACCATTCCCAGCTAATTTTTGTATTTTTGGTAGAGACAGGGTTTCA
TCATGTTATCCAGGCTGGTCTCAAACTCCTGGACTCAAATAATCCACCAGCCTTGGCCTC
CCAAAGTGCTGGGATTAAAGGCATGAGCCACTGCGCCCGGCCATGCATGTTTTCTTTCTT
GCCTGGTAGGCAGGAATCTCTCAACTTATTTCTGACTTTCTCTCACAGGGAATTAATTGA
GATGTTCATTCTGTGCATTTGTGAGTATTGGGAGTGCCAGGAGCTTCCTATTCTGCCATG
TTGCTGACATCAGTCTAAGGAAAACAGTTTAAAGAAAGTTCATCAAAAAGTAACAGTAGA
CACATCTGGGTGTCTTAAATATGAATACATTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT
CTTTCTTTCTTTCCTTCTTTTCTTTCTTCCTTTCC

FIG. 1AY